



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 120900

TO: Jeffrey Parkin
Location: rem/3d39/3c18
Art Unit: 1648
Monday, May 03, 2004

Case Serial Number: 09/775964

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Parkin,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

120900

From: Parkin, Jeffrey
Sent: Sunday, May 02, 2004 2:02 PM
To: STIC-Biotech/ChemLib
Subject: U.S. Serial No. 09/775,964

Please search the following SEQ ID NOS. v. all relevant databases (including interference): SEQ ID NOS.: 5 and 26.

Place results on both paper and disk.

Thanks!

JSP
Au 1648
REM 3D39
2-0908

RECEIVED
MAY - 3 2004
STIC/CHEN, Division

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 5/3/04
Date Completed: 5/3/04
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: ✓
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: ASP
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:05:13 ; Search time 72.4946 Seconds
(without alignments)

1781.157 Million cell updates/sec

Title: US-09-775-964-5

Perfect score: 2398

Sequence: 1 PDLRFTNIGPDMRTVWAP.....LVTLPHNLGHPILDPVST 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1980s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2398	100.0	457	2 AAW33340	Aaw33340 Human fib
2	2257	94.1	432	2 AAR40160	Aar40160 Human FN/
3	2257	94.1	432	2 AAW33339	Aaw33339 Human fib
4	2257	94.1	432	2 AAY05457	Aay05457 Fibronect
5	2257	94.1	432	6 ABP72002	Abp72002 Human fib
6	2257	94.1	455	6 ABP72004	Abp72004 Human fib
7	2245	93.6	432	2 AAR40162	Aar40162 Human FN/
8	1536.5	64.1	574	2 AAR08044	Aar08044 Fibronect
9	1536.5	64.1	574	2 AAR60351	Aar60351 Inhibitor
10	1536.5	64.1	574	2 AAW33349	Aaw33349 Oligopept
11	1536.5	64.1	574	2 AAW13572	Aaw13572 Escherich
12	1536.5	64.1	574	2 AAY05453	Aay05453 Fibronect
13	1536.5	64.1	574	2 AAW97357	Aaw97357 Protein s
14	1536.5	64.1	574	6 ABP59436	Abp59436 Human fib
15	1536.5	64.1	574	7 ADD49015	Add49015 Fibronect
16	1527	63.7	489	2 AAW33343	Aaw33343 Protein u
17	1523	63.5	573	2 AAR60350	Aar60350 Chimeric
18	1501	62.6	2265	4 AAM38647	Aam38647 Human pol
19	1485.5	61.9	302	2 AAR60355	Aar60355 Chimeric
20	1485.5	61.9	302	2 AAR98816	Aar98816 Cell adhe
21	1485.5	61.9	302	2 AAW33352	Aaw33352 Oligopept
22	1485.5	61.9	302	6 ABP59437	Abp59437 Human fib
23	1485.5	61.9	302	7 ADD49016	Add49016 Fibronect
24	1474	61.5	2320	6 AAE37107	Aae37107 Human fib
25	1474	61.5	2320	6 ABR42588	Abr42588 Human fib

ALIGNMENTS

RESULT 1

AAW33340

ID AAW33340 standard; protein; 457 AA.

AC AAW33340;

XX

DT 23-FEB-1998 (first entry)

XX

DE Human fibronectin fragment containing CS-I binding domain.

XX

XX

KW CS-I binding domain; human fibronectin; target cell; transfection;

KW retroviral vector; gene therapy; cancer; viral disease;

KW acquired immunodeficiency syndrome; AIDS.

XX

OS Homo sapiens.

XX

PN WO9718318-A1.

XX

PD 22-MAY-1997.

XX

PF 07-NOV-1996; 96WO-JP003254.

XX

PR 13-NOV-1995; 95JP-00294382.

PR 08-MAR-1996; 96JP-00051847.

XX

XX (TAKI) TAKARA SHUZO CO LTD.

XX

PI Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;

XX

XX WPI; 1997-289294/26.

DR N-PSDB; AAT93970.

XX

PT Method for increasing efficacy of gene transfer to target cell using

PT retrovirus - by infection of the target cell in the presence of a

PT substance which binds to the virus and a substance which binds to the

PT target cell.

XX

PS Claim 42; Page 97-99; 194pp; Japanese.

XX

CC The present sequence is a human fibronectin fragment containing a CS-I

CC binding domain, which was used in the development of a novel method for

CC increasing the efficiency of gene introduction into a target cell using a

CC retroviral vector. The method comprises carrying out viral infection of

CC the target cell in the presence of a retrovirus and target cell binding

CC substance or substances. The method can be used to effectively introduce

CC genes into target cells for the gene therapy of cancer and viral

CC diseases, e.g. AIDS

XX

Abo01288 Human pro
Aar92778 Human fib
Aau74674 Human fib
Aae23651 Human pro
Aap70373 Human fib
Aar15468 Human fib
Aag68182 Fibronect
Abu07486 Protein d
Abr41106 Human fib
Abr92078 Human cer
Adb70378 Fibronect
Adb98726 Human fib
Ade82522 Human pro
Aaw58335 NM_00202
Aaw63171 Amino aci
Aao17353 Human fib
Aar81866 Human fib
Add18770 Human dis
Ade63324 Human Pro
Aar60021 Fibrinoge

SQ Sequence 457 AA;
Query Match 100.0%; Score 2398; DB 2; Length 457;
Best Local Similarity 100.0%; Pred. No. 3.1e-172;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PTDLRFTNIGDTRVTVWAPPSSIDLTFNVLRYSPVKNEEDVAELSISSPDNAVLTNLL 60
DB 1 PTDLRFTNIGDTRVTVWAPPSSIDLTFNVLRYSPVKNEEDVAELSISSPDNAVLTNLL 60
QY 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDSDPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDSDPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
QY 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGGSGAPPP 300
DB 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGGSGAPPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIGVCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDECFFERLESNNYRYSRKYTSWYVALKRTGYKLGSKTGPQ 420
DB 361 AMKEDGRLLASKCVTDECFFERLESNNYRYSRKYTSWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAASDELQVLTLPHPNLHGPEILDVPST 457
DB 421 KAILFLPMSAASDELQVLTLPHPNLHGPEILDVPST 457
RESULT 2
AAR40160
ID AAR40160 standard; peptide; 432 AA.
AC AAR40160;
XX 07-FEB-1994 (first entry)
DE Human FN/bFGF fusion peptide #1.
XX Human; fibronectin; FN; fibroblast cell growth factor; FGF; fusion;
KW cell adhesion; cell growth; anti-aging; cosmetics; wound healing;
KW surgery.
XX Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..277
FT Peptide /note= "Human FN fragment 1239-1515"
FT Peptide 278..432
FT Peptide /note= "bFGF fragment"
XX
PN JP05178897-A.
XX 20-JUL-1993.
XX
XX 05-MAR-1992; 92JP-00083220.
XX
XX 14-OCT-1991; 91JP-00291959.
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX WPI; 1993-261656/33.
XX

PT Synthetic functional polypeptide to promote wound healing, etc. - contg.
PT cell adhesion polypeptide from fibronectin and fibroblast growth factor
PT polypeptide, opt. linked by spacer.
XX Disclosure; Page 7-8; 13pp; Japanese.
XX The sequences given in AAR40158-63 represent human fibronectin (FN) and
CC fibroblast cell growth factor (FGF) fragments which were used in the
CC production of fusion polypeptides which are able to stimulate cell
CC adhesion and cell growth. these fusion peptides may be used for anti-
CC aging cosmetics and in wound healing after surgery
XX Sequence 432 AA;
SQ Query Match 94.1%; Score 2257; DB 2; Length 432;
Best Local Similarity 99.8%; Pred. No. 1.2e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PTDLRFTNIGDTRVTVWAPPSSIDLTFNVLRYSPVKNEEDVAELSISSPDNAVLTNLL 60
DB 1 PTDLRFTNIGDTRVTVWAPPSSIDLTFNVLRYSPVKNEEDVAELSISSPDNAVLTNLL 60
QY 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDSDPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDSDPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
QY 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGGSGAPPP 300
DB 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGGSGAPPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIGVCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDECFFERLESNNYRYSRKYTSWYVALKRTGYKLGSKTGPQ 420
DB 361 AMKEDGRLLASKCVTDECFFERLESNNYRYSRKYTSWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
DB 421 KAILFLPMSAKS 432
RESULT 3
AAW33339
ID AAW33339 standard; protein; 432 AA.
AC AAW33339;
XX 23-FEB-1998 (first entry)
DT Human fibroblast growth factor oligopeptide.
XX Human; fibroblast growth factor; FGF; target cell; transfection;
KW retroviral vector; gene therapy; cancer; viral disease;
KW acquired immunodeficiency syndrome; AIDS.
XX Homo sapiens.
OS
XX WO9718318-A1.
XX 22-MAY-1997.
XX 07-NOV-1996; 96WO-JP003254.
XX

PR 13-NOV-1995; 95JP-00294382.
PR 08-MAR-1996; 96JP-00051847.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;
XX
XX WPI; 1997-289294/26.
DR
XX
XX Method for increasing efficacy of gene transfer to target cell using
PT retrovirus - by infection of the target cell in the presence of a
PT substance which binds to the virus and a substance which binds to the
PT target cell.
XX
XX Claim 42; Page 94-97; 194pp; Japanese.
PS
XX
XX The present sequence is a human fibroblast growth factor (FGF)
CC oligopeptide, which was used in the development of a novel method for
CC increasing the efficiency of gene introduction into a target cell using a
CC retroviral vector. The method comprises carrying out viral infection of
CC the target cell in the presence of a retrovirus and target cell binding
CC substance or substances. The method can be used to effectively introduce
CC genes into target cells for the gene therapy of cancer and viral
CC diseases, e.g. AIDS
XX
SQ Sequence 432 AA;
Query Match 94.1%; Score 2257; DB 2; Length 432;
Best Local Similarity 99.8%; Pred. No. 1.2e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTNFLVRYSPVKNEEDVAELSIKSPSDNAVLTNLL 60
DB 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTNFLVRYSPVKNEEDVAELSIKSPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFGSRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFGSRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGGAFPP 300
DB 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGGAFPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAERGVSIIKGVCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAERGVSIIKGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDECFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQG 420
DB 361 AMKEDGRLLASKCVTDECFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQG 420
QY 421 KAILFLPMSAAS 432
DB 421 KAILFLPMSAKS 432
RESULT 4
AA05457
ID RAY05457 standard; protein; 432 AA.
XX
AC AA05457;
XX
DT 07-JUL-1999 (first entry)
XX
DE Fibronectin receptor targeting HIV strain C-FGF.A.

XX Fibronectin receptor; HIV; infection; therapy.
XX
XX Unidentified.
XX JP10029952-A.
XX
XX 03-FEB-1998.
XX
XX 16-JUL-1996; 96JP-00185893.
XX
XX 16-JUL-1996; 96JP-00185893.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX WPI; 1998-163674/15.
XX
XX Control of human immunodeficiency virus infection - using composition
PT comprising replication defective HIV vector.
XX
XX Example 3; Page 21-22; 24pp; Japanese.
XX
XX This sequence represents a fibronectin receptor that can be used in the
CC method of the invention. The method is for the control of human
CC immunodeficiency virus (HIV) infection using a composition which
CC comprises a functional substance which participates in the infection of
CC HIV. The method is used to control HIV-infection
XX
SQ Sequence 432 AA;
Query Match 94.1%; Score 2257; DB 2; Length 432;
Best Local Similarity 99.8%; Pred. No. 1.2e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTNFLVRYSPVKNEEDVAELSIKSPSDNAVLTNLL 60
DB 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTNFLVRYSPVKNEEDVAELSIKSPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFGSRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFGSRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGGAFPP 300
DB 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGGAFPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAERGVSIIKGVCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAERGVSIIKGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDECFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQG 420
DB 361 AMKEDGRLLASKCVTDECFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQG 420
QY 421 KAILFLPMSAAS 432
DB 421 KAILFLPMSAKS 432
RESULT 5
ABP72002
ID ABP72002 standard; protein; 432 AA.
XX
XX ABP72002;
XX

DT 16-MAY-2003 (first entry)
XX Human fibroblast growth factor C-FGF.
DE
KW Human; drug; cell adhesion; fibroblast growth factor; FGF; cardiant;
KW osteopathic; angiogenesis promotion; arterial disease; cartilage damage;
KW C-FGF.
XX
OS Homo sapiens.
OS Synthetic.
XX WO2003004066-A1.
XX
XX 16-JAN-2003.
XX
XX 04-JUL-2002; 2002WO-JP006772.
XX
XX 05-JUL-2001; 2001JP-00204300.
XX
XX (TAKA-) TAKARA BIO INC.
XX Tanaka K, Chono H, Morishita M, Konishi H, Ueno M, Asada K;
XX Kato I;
XX
XX WPI; 2003-210320/20.
XX N-PSDB; ABZ82340.
XX
XX Gene therapeutics to provide fusion polypeptide with cell adhesion and
XX fibroblast growth factor activities, applicable in angiogenesis promotion
XX and treatment of arterial diseases and cartilage damage.
XX
XX Claim 8; Page 36-38; 53pp; Japanese.
XX
XX The invention relates to novel drugs containing a gene encoding a
XX polypeptide which has the amino acid sequence of a polypeptide having a
XX cell adhesion activity and another polypeptide with an amino acid
XX sequence of a polypeptide having a fibroblast growth factor activity. The
XX drugs of the invention have cardiant, and osteopathic activity. The gene
XX therapeutics of the invention are applicable in angiogenesis promotion
XX and treatment of arterial diseases and cartilage damage. The sequence
XX represents the human C-FGF protein, a fusion of C277 and bFGF
XX
SQ Sequence 432 AA;

Query Match 94.1%; Score 2257; DB 6; Length 432;
Best Local Similarity 99.8%; Pred. No. 1.2e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTDLRFNIGPDTMRVTWAPPSSIDLTFNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLRFNIGPDTMRVTWAPPSSIDLTFNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60

QY 61 PGTEYVSVSSVYQHESTPLRGOKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 120
DB 61 PGTEYVSVSSVYQHESTPLRGOKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 120

QY 121 IRRHPEHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQSTVSD 180
DB 121 IRRHPEHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQSTVSD 180

QY 181 VPRDLEVAATTSLLISWDAPAVTVRYRIYTGTSNPSVQVETVPGSKSTATISGLK 240
DB 181 VPRDLEVAATTSLLISWDAPAVTVRYRIYTGTSNPSVQVETVPGSKSTATISGLK 240

QY 241 PGVDYITIVYATGSDSPASSKPIINVRTEIDKPSMAAGSITTLPALPEDGGGAPPP 300
DB 241 PGVDYITIVYATGSDSPASSKPIINVRTEIDKPSMAAGSITTLPALPEDGGGAPPP 300

QY 301 GHFKDPKRLYCKNGGFFLRHDPGRVDGVREKSDPHIKLQQAERGVVSIKGVCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHDPGRVDGVREKSDPHIKLQQAERGVVSIKGVCANRYL 360

QY 361 AMKEDGRLLASKCVTDECFERLESNNYVRSRKYTSWYVALKRTGYKLGSKTGPGQ 420

DB 361 AMKEDGRLLASKCVTDECFERLESNNYVRSRKYTSWYVALKRTGYKLGSKTGPGQ 420
QY 421 KAILFLPMSAAS 432
DB 421 KAILFLPMSAKS 432

RESULT 6
ABP72004
ID ABP72004 standard; protein; 455 AA.
XX
XX AC ABP72004;
XX
XX 16-MAY-2003 (first entry)
XX
XX Human fibroblast growth factor SC-FGF.
DE Human; drug; cell adhesion; fibroblast growth factor; FGF; cardiant;
KW osteopathic; angiogenesis promotion; arterial disease; cartilage damage;
KW SC-FGF.
XX
XX Homo sapiens.
XX
XX WO2003004066-A1.
XX
XX 16-JAN-2003.
XX
XX 04-JUL-2002; 2002WO-JP006772.
XX
XX 05-JUL-2001; 2001JP-00204300.
XX
XX (TAKA-) TAKARA BIO INC.
XX Tanaka K, Chono H, Morishita M, Konishi H, Ueno M, Asada K;
XX Kato I;
XX
XX WPI; 2003-210320/20.
XX N-PSDB; ABZ82342.
XX
XX Gene therapeutics to provide fusion polypeptide with cell adhesion and
XX fibroblast growth factor activities, applicable in angiogenesis promotion
XX and treatment of arterial diseases and cartilage damage.
XX
XX Claim 13; Page 41-43; 53pp; Japanese.
XX
XX The invention relates to novel drugs containing a gene encoding a
XX polypeptide which has the amino acid sequence of a polypeptide having a
XX cell adhesion activity and another polypeptide with an amino acid
XX sequence of a polypeptide having a fibroblast growth factor activity. The
XX drugs of the invention have cardiant, and osteopathic activity. The gene
XX therapeutics of the invention are applicable in angiogenesis promotion
XX and treatment of arterial diseases and cartilage damage. The sequence
XX represents the human SC-FGF protein, a fusion of FGF-4 signal and C-FGF
XX
SQ Sequence 455 AA;

Query Match 94.1%; Score 2257; DB 6; Length 455;
Best Local Similarity 99.8%; Pred. No. 1.3e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTDLRFNIGPDTMRVTWAPPSSIDLTFNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 24 PTDLRFNIGPDTMRVTWAPPSSIDLTFNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 83

QY 61 PGTEYVSVSSVYQHESTPLRGOKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 120
DB 84 PGTEYVSVSSVYQHESTPLRGOKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 143

QY 121 IRRHPEHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQSTVSD 180
DB 144 IRRHPEHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQSTVSD 203

QY 181 VPRDLVVAATPTSLISWDAPAVTVYRITYGETGNSPVQEFVPGSKSTATISGLK 240
Db 204 VPRDLVVAATPTSLISWDAPAVTVYRITYGETGNSPVQEFVPGSKSTATISGLK 263
QY 241 PGVDYTTITVAVTGRGDSPASSKPI SINRYRTEDKPSMAAGSIITLTPALPEDGSGGAFPP 300
Db 264 PGVDYTTITVAVTGRGDSPASSKPI SINRYRTEDKPSMAAGSIITLTPALPEDGSGGAFPP 323
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGWSIKGVCANRYL 360
Db 324 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGWSIKGVCANRYL 383
QY 361 AMKEDGRLASKCVTDECFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQ 420
Db 384 AMKEDGRLASKCVTDECFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQ 443
QY 421 KAILFLPMSAAS 432
Db 444 KAILFLPMSAKS 455

RESULT 7
AAR40162
ID AAR40162 standard; peptide; 432 AA.
XX
AC AAR40162;
XX
DT 07-FEB-1994 (first entry)
XX
DE Human FN/bFGF fusion peptide #2.
XX
KW Human; fibronectin; FN; fibroblast cell growth factor; FGF; fusion;
KW cell adhesion; cell growth; anti-aging; cosmetics; wound healing;
KW surgery.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..277
FT /note= "Human FN fragment 1239-1515"
FT Peptide 278..432
FT /note= "bFGF fragment"
XX
PN JP05178897-A.
XX
PD 20-JUL-1993.
XX
PF 05-MAR-1992; 92JP-00083220.
XX
PR 14-OCT-1991; 91JP-00291959.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
DR WPI; 1993-261656/33.
XX
FT Synthetic functional polypeptide to promote wound healing, etc. - contg.
FT cell adhesion polypeptide from fibronectin and fibroblast growth factor
FT polypeptide, opt. linked by spacer.
XX
PS Disclosure; Page 9-10; 13pp; Japanese.
XX
CC The sequences given in AAR40158-63 represent human fibronectin (FN) and
CC fibroblast cell growth factor (FGF) fragments which were used in the
CC production of fusion polypeptides which are able to stimulate cell
CC adhesion and cell growth. These fusion peptides may be used for anti-
CC aging cosmetics and in wound healing after surgery
XX
SQ Sequence 432 AA;

Query Match 93.6%; Score 2245; DB 2; Length 432;
Best Local Similarity 99.3%; Pred. No. 9.8e-161;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PTDLRTNIGPDMRTWAPPSSIDLTLNLFVRSYVKNEDVAELSISSPSDNAVLTLNLL 60
Db 1 PTDLRTNIGPDMRTWAPPSSIDLTLNLFVRSYVKNEDVAELSISSPSDNAVLTLNLL 60
QY 61 PGTEYVSVSSVVEQHESTPLRGROKTLGDSPTGIDFSDITANSFTVHWTAPRATTIGYR 120
Db 61 PGTEYVSVSSVVEQHESTPLRGROKTLGDSPTGIDFSDITANSFTVHWTAPRATTIGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVISVALNGREESPLLIQQOSTVSD 180
Db 121 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVISVALNGREESPLLIQQOSTVSD 180
QY 181 VPRDLVVAATPTSLISWDAPAVTVYRITYGETGNSPVQEFVPGSKSTATISGLK 240
Db 181 VPRDLVVAATPTSLISWDAPAVTVYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTTITVAVTGRGDSPASSKPI SINRYRTEDKPSMAAGSIITLTPALPEDGSGGAFPP 300
Db 241 PGVDYTTITVAVTGRGDSPASSKPI SINRYRTEDKPSMAAGSIITLTPALPEDGSGGAFPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGWSIKGVCANRYL 360
Db 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGWSIKGVCANRYL 360
QY 361 AMKEDGRLASKCVTDECFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQ 420
Db 361 AMKEDGRLASKCVTDECFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
Db 421 KAILFLPMSAKS 432

RESULT 8
AAR08044
ID AAR08044 standard; protein; 574 AA.
XX
AC AAR08044;
XX
DT 25-MAR-2003 (revised)
DT 25-FEB-1991 (first entry)
XX
DE Fibronectin deriv. for inhibition of angiogenesis.
XX
KW Fibronectin; heparin; angiogenesis; metastasis; tumours; prolasias;
KW contraceptive; retinopathy; osteoporosis; rheumatism.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..277
FT /label= FN fragment of cell-binding domain
FT Peptide 279..549
FT /label= FN fragment of heparin-binding domain
XX
PN EP399806-A.
PD 28-NOV-1990.
XX
PF 23-MAY-1990; 90EP-00305619.
XX
PR 26-MAY-1989; 89JP-00131453.
PR 01-DEC-1989; 89JP-00310536.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Taguchi Y, Ohdate Y, Kawase Y, Goto S, Fusao K, Kato I, Saiki I;
PI Azuma I;
XX
WPI; 1990-356419/48.
XX
DE Fibronectin deriv. - contg. cell binding and heparin-binding domains
XX linked via aminoacid residue or peptide, used as inhibitor of


```
QY 399 -----SWYVALKRTGQY-----KLGSK-----TG- 417
Db 462 DAPSNLRLATPNSLLVSWQPPRARITGVIKYEKGSPPREVPRPGVTEATITGL 521
QY 418 -PGOKAILFL-----PM--SAASDELPLQVTLPHNPHNLHGPEILDVPST 457
Db 522 EPGTEYTIYVIALKNQKSEPLIGRKKTDLPQLVTLPHNPHNLHGPEILDVPST 574

RESULT 10
AAW33349
ID AAW33349 standard; protein; 574 AA.
XX
AC AAW33349;
XX
DT 23-FEB-1998 (first entry)
XX
DE Oligopeptide CH-296.
XX
KW Oligopeptide CH-296; target cell; transfection; retroviral vector;
KW gene therapy; cancer; viral disease; acquired immunodeficiency syndrome;
KW AIDS.
XX
OS Synthetic.
XX
PN WO9718318-A1.
XX
PD 22-MAY-1997.
XX
PF 07-NOV-1996; 96WO-JP003254.
XX
PR 13-NOV-1995; 95JP-00294382.
PR 08-MAR-1996; 96JP-00051847.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;
XX
DR WPI; 1997-289294/26.
XX
PT Method for increasing efficacy of gene transfer to target cell using
PT retrovirus - by infection of the target cell in the presence of a
PT substance which binds to the virus and a substance which binds to the
PT target cell.
XX
PS Disclosure; Page 130-133; 194pp; Japanese.
XX
CC The present sequence is the oligopeptide CH-296, which was used in the
CC development of a novel method for increasing the efficiency of gene
CC introduction into a target cell using a retroviral vector. The method
CC comprises carrying out viral infection of the target cell in the presence
CC of a retrovirus and target cell binding substance or substances. The
CC method can be used to effectively introduce genes into target cells for
CC the gene therapy of cancer and viral diseases, e.g. AIDS
XX
SQ Sequence 574 AA;

Query Match 64.1%; Score 1536.5; DB 2; Length 574;
Best Local Similarity 58.0%; Pred. No. 3.5e-107;
Matches 34%; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PTDLRFNIGDTRVMTWAPPSPSIDLTNFLVRYSPVKNEEDVAELISPSDNAVLTNLL 60
Db 1 PTDLRFNIGDTRVMTWAPPSPSIDLTNFLVRYSPVKNEEDVAELISPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDPSPTGIDFSDITANSFTVHWIAPRATITGYR 120
Db 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDPSPTGIDFSDITANSFTVHWIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGRESPLLIQQQSTVSD 180
Db 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGRESPLLIQQQSTVSD 180
```

```
QY 181 VPRDLEVAATPTSLILISWDAPAVTVYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLILISWDAPAVTVYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYTIIVYAVTGRGDSPASSKPIISINVRTEIDKPSMAAGSITTL---PALPEDGSGA 297
Db 241 PGVDYTIIVYAVTGRGDSPASSKPIISINVRTEIDKPSMAIPAPTDLKFQVTTPTSLSAQW 300
QY 298 FPEGHPKDKRLYCKNGGPFRLIHPDGRVDGVREKSDPHIKLQQAEEGVVSIKGYCAN 357
Db 301 TPN-----VLTGYRVVTP-----KEKTGPMKEINLAPDSSSVVWVGLMVA 344
QY 358 RY-----LANKEDGRLLASK-----CVTDEC----- 378
Db 345 KYEVSIVALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETTTISWRTKTETITG 401
QY 379 FPFERLESNNYNTYRS-----RKYT----- 398
Db 402 FQVDVAPANGQTPIQRTIKPDVRSYITGLQPGDYKIYLYTLNDNARSFVIDASTAI 461
QY 399 -----SWYVALKRTGQY-----KLGSK-----TG- 417
Db 462 DAPSNLRLATPNSLLVSWQPPRARITGVIKYEKGSPPREVPRPGVTEATITGL 521
QY 418 -PGOKAILFL-----PM--SAASDELPLQVTLPHNPHNLHGPEILDVPST 457
Db 522 EPGTEYTIYVIALKNQKSEPLIGRKKTDLPQLVTLPHNPHNLHGPEILDVPST 574

RESULT 11
AAW13572
ID AAW13572 standard; protein; 574 AA.
XX
AC AAW13572;
XX
DT 14-NOV-1997 (first entry)
XX
DE Escherichia coli fibronectin protein.
XX
KW Immunology; haematopoietic somatic cell; vertebrate; mammal; bird;
KW amphibian; fish; reptile; mouse; human; bone marrow.
XX
OS Escherichia coli.
XX
PN WO9707670-A1.
XX
PD 06-MAR-1997.
XX
PF 19-AUG-1996; 96WO-JP002312.
XX
PR 30-AUG-1995; 95JP-00221583.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Matsushita H, Kato I;
XX
DR WPI; 1997-178821/16.
XX
PT Vertebrate immunologically tolerant to foreign proteins - produced by
PT transforming haematopoietic somatic cells with a gene encoding foreign
PT protein.
XX
PS Disclosure; Page 34-37; 48pp; Japanese.
XX
CC A method has been developed for producing a vertebrate immunologically
CC tolerant to foreign proteins. The method involves: (a) transforming
CC haematopoietic somatic cells of a vertebrate with a gene encoding foreign
CC proteins, (b) producing an immunodeficient vertebrate, and (c)
CC transferring the transformed cells into the immunodeficient vertebrate.
CC The present sequence represents an Escherichia coli fibronectin protein
CC which can be used as an example of the above method. The vertebrate
CC may be used in the production of foreign proteins with pharmaceutical
CC applications. They may also have potential uses as models for human
```

CC diseases
XX
SQ Sequence 574 AA;

Query Match 64.1%; Score 1536.5; DB 2; Length 574;
Best Local Similarity 58.0%; Pred. No. 3.5e-107;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PTDLRFNIGPDTMRVTWAPPSSIDLTNPLVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLRFNIGPDTMRVTWAPPSSIDLTNPLVRYSPVKNEEDVAELISPSDNAVLTNLL 60

QY 61 PGTYYVSVSSVYEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTYYVSVSSVYEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 180

QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240

QY 241 PGVDYITIVYAVTGRGDSPASSKPI SINRYRTEIDKPSMAAGSITTL---PALPEDGSGA 297
DB 241 PGVDYITIVYAVTGRGDSPASSKPI SINRYRTEIDKPSMAIPAPTDLKTQVTPTSLSAQW 300

QY 298 FPPGHFKDPKRLCKNGGFFLRHPDGRVDGVREKSDPHIKLOLQAEERGVSIVKGVCAN 357
DB 301 TPNP-----VQUTGYRVVTP-----KEKTGPMKEINLAPDSSVVVSGLMVAT 344

QY 358 RY----LAMKEDGRLLASK-----CVTDEC----- 378
DB 345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPPRARVTDATETITISWRTKTETITG 401

QY 379 FFERLESNNYTRS-----RKYT----- 398
DB 402 FQVDAVPANGQTPQRTIKPDVRSYITITGLQGTDYKIYLYTLNDNARSSPVVIDASTAI 461

QY 399 -----SWYVALKRTGOY-----KLGSK-----TG- 417
DB 462 DAPSNLRFATPNSLLVSWQPPRARIITYIKYKPGSPPREVVRPRPGVTEATITGL 521

QY 418 -PGQKAILFL-----PM--SAASDELQVLTPLPHNHLGPEILDVPST 457
DB 522 EPGTEYTIIVIALKNNQKSEPLIGRKKTDDELQVLTPLPHNHLGPEILDVPST 574

RESULT 12
AA05453
ID AAY05453 standard; protein; 574 AA.
AC AAY05453;
XX
XX
DT 07-JUL-1999 (first entry)
XX
XX Fibronectin receptor.
XX Fibronectin receptor; HIV; infection; therapy.
XX Unidentified.
XX JP10029952-A.
XX
XX 03-FEB-1998.
XX
XX 16-JUL-1996; 96JP-00185893.
XX
XX 16-JUL-1996; 96JP-00185893.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX

DR WPI; 1998-163674/15.
XX Control of human immunodeficiency virus infection - using composition
PT comprising replication defective HIV vector.
XX
XX Claim 3; Page 11-14; 24pp; Japanese.
XX
CC This sequence represents a fibronectin receptor that can be used in the
CC method of the invention. The method is for the control of human
CC immunodeficiency virus (HIV) infection using a composition which
CC comprises a functional substance which participates in the infection of
CC HIV. The method is used to control HIV-infection
XX
SQ Sequence 574 AA;

Query Match 64.1%; Score 1536.5; DB 2; Length 574;
Best Local Similarity 58.0%; Pred. No. 3.5e-107;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PTDLRFNIGPDTMRVTWAPPSSIDLTNPLVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLRFNIGPDTMRVTWAPPSSIDLTNPLVRYSPVKNEEDVAELISPSDNAVLTNLL 60

QY 61 PGTYYVSVSSVYEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTYYVSVSSVYEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 180

QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240

QY 241 PGVDYITIVYAVTGRGDSPASSKPI SINRYRTEIDKPSMAAGSITTL---PALPEDGSGA 297
DB 241 PGVDYITIVYAVTGRGDSPASSKPI SINRYRTEIDKPSMAIPAPTDLKTQVTPTSLSAQW 300

QY 298 FPPGHFKDPKRLCKNGGFFLRHPDGRVDGVREKSDPHIKLOLQAEERGVSIVKGVCAN 357
DB 301 TPNP-----VQUTGYRVVTP-----KEKTGPMKEINLAPDSSVVVSGLMVAT 344

QY 358 RY----LAMKEDGRLLASK-----CVTDEC----- 378
DB 345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPPRARVTDATETITISWRTKTETITG 401

QY 379 FFERLESNNYTRS-----RKYT----- 398
DB 402 FQVDAVPANGQTPQRTIKPDVRSYITITGLQGTDYKIYLYTLNDNARSSPVVIDASTAI 461

QY 399 -----SWYVALKRTGOY-----KLGSK-----TG- 417
DB 462 DAPSNLRFATPNSLLVSWQPPRARIITYIKYKPGSPPREVVRPRPGVTEATITGL 521

QY 418 -PGQKAILFL-----PM--SAASDELQVLTPLPHNHLGPEILDVPST 457
DB 522 EPGTEYTIIVIALKNNQKSEPLIGRKKTDDELQVLTPLPHNHLGPEILDVPST 574

RESULT 13
AAW97357
ID AAW97357 standard; peptide; 574 AA.
XX
XX AAW97357;
XX
XX 12-MAY-1999 (first entry)
XX
XX Protein sequence of the specification.
XX Retrovirus; gene transfer; serum-free medium; AIDS; cancer; leukaemia;
KW gene therapy.
XX

OS Homo sapiens.
FN WO9905301-A1.
XX 04-FEB-1999.
XX 15-JUL-1998; 98WO-JP003173.
XX 23-JUL-1997; 97JP-00196772.
XX (TAKI) TAKARA SHUZO CO LTD.
XX Bagnis C, Imbert A, Mannoni P;
XX WPI; 1999-142951/12.
XX
XX Gene transfer by retrovirus in medium containing functional substance and
XX optionally low-density lipoprotein - useful in medical sciences, cell and
XX gene engineering, particularly for treating AIDS and cancers.
XX
XX Claim 5; Page 27-30; 32pp; Japanese.
XX
XX The specification describes a method for transferring a gene into target
XX cells by a retrovirus using a serum-free medium. The culture medium of
XX the target cells is serum free and contains an effective amount of a
XX functional substance to elevate the gene transfer efficiency when both
XX the retrovirus and target cells are present together. The gene transfer
XX method is useful in medical sciences, cell engineering and genetic
XX engineering, such as in the treatment of AIDS and cancers e.g. leukaemia
XX by gene therapy
XX
SQ Sequence 574 AA;

Query Match 64.1%; Score 1536.5; DB 2; Length 574;
Best Local Similarity 58.0%; Pred. No. 3.5e-107;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;
QY 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTLNLFVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60
DB 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTLNLFVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTLNTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTLNTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITTVYAVTGRGDSPPASSKPIISINYTEIDKPSMAIGSITTL---PALPEDGSGA 297
DB 241 PGVDYITTVYAVTGRGDSPPASSKPIISINYTEIDKPSMAIPAPTDLKTQVTFSLAOW 300
QY 298 FPGCHFKDPKRLCYCKNGFFLRHPDGRVDGVRKSDPHIKLOLQAEERGWWSIKGCAN 357
DB 301 TPN-----VOLTGYRVTP-----KEKTPMKKEINLAPDSSVVVSGLMVAT 344
QY 358 RY----LAMKEDGRLLASK-----CVTDEC----- 378
DB 345 KYEVSVALKD---TLTSRAQGVVTLLENVSPRRARVTDATETTITISWRKTTETIG 401
QY 379 FFERLESNNYTRS-----RKYT----- 398
DB 402 FOYDAVPANGQTPQRTIKPDVRSYITLQPGQDYKIVLYTLNDNARSPVVIDASTAI 461
QY 399 -----SNVVALKRTGY-----KLGSK-----TG- 417
DB 462 DAFSNRFLATTNSLLVSHQPPRAITGYIIKYEKGGSPREVPRPVGVTETATIGL 521

QY 418 -PCQKAILFL-----PM--SAASDELPLQVTLPHNHLHGPEILDVPST 457
DB 522 EPGTEYTIYVIALKNQKSEPLIGRKTKDELPLQVTLPHNHLHGPEILDVPST 574
RESULT 14
ABP59436
ID ABP59436 standard; protein; 574 AA.
XX AC ABP59436;
XX 09-JUN-2003 (first entry)
XX Human fibronectin fragment CH-296, SEQ ID 6.
XX Human; fibronectin; antibacterial; virucide; cytostatic; antiallergic;
XX cytotoxic T-cell; CTL; cytotoxicity; CD44; cancer; infection.
XX Homo sapiens.
XX WO2003016511-A1.
XX 27-FEB-2003.
XX 15-AUG-2002; 2002WO-JP008298.
XX 15-AUG-2001; 2001JP-00246747.
XX 11-DEC-2001; 2001JP-00376966.
XX 25-MAR-2002; 2002JP-00084428.
XX (TAKA-) TAKARA BIO INC.
XX Sagawa H, Ideno M, Kato I;
XX WPI; 2003-289979/28.
XX Induction of antigen-specific cytotoxic T cells by culture in presence of
XX a substance binding to CD44 for production of therapeutic cells treating
XX infectious and allergic disorders and cancer.
XX Disclosure; Page 102-106; 142pp; Japanese.
XX
XX The present invention relates to a method for inducing cytotoxic T-cells
XX (CTL) having antigen-specific cytotoxicity. The method comprises
XX incubating cells capable of differentiating into CTL with a cellular or
XX non-cellular antigen in the presence of a substance having one of the
XX following properties: (1) binding to CD44; (2) regulating the signal
XX generated by CD44 binding to its ligand; (3) inhibiting binding of a
XX growth factor to its receptor; (4) regulating the signal generated by
XX growth factor binding to its receptor; (5) containing fibronectin and/or
XX its fragments. Antigen-specific CTL are used for the treatment of cancer
XX and bacterial and viral infections. The present sequence is a fragment of
XX human fibronectin, which was used to illustrate the method of the
XX invention
XX
SQ Sequence 574 AA;

Query Match 64.1%; Score 1536.5; DB 6; Length 574;
Best Local Similarity 58.0%; Pred. No. 3.5e-107;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;
QY 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTLNLFVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60
DB 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTLNLFVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTLNTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTLNTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180

QY 181 VPRDLEVAATPTSLLSWDAVAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLLSWDAVAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTIIVYAVTGRGDSPASSKPIINRYTEIDKPSMAAGSITTL---PALPEDGSGA 297
Db 241 PGVDYTIIVYAVTGRGDSPASSKPIINRYTEIDKPSMAIPAPTDLKTQVTPTSLSAQW 300
QY 298 FPGGHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAEEGVVSIKGVCAN 357
Db 301 TPN-----VLTGYRVVTP-----KEKTPMKEINLAPDSSVVVSGLMVAT 344
QY 358 RY----LAWKEDGRLLASK-----CVTDEC----- 378
Db 345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETITISWRTKTETITG 401
QY 379 FFERLESNNYTYRS-----RKYT----- 398
Db 402 FOVDVAVPANGOTPIORTIKPDVRSYITGLQPGTDYKIYLYTLNDNARSPVVIDASTAI 461
QY 399 -----SWYVALKRTGOY-----KLGSK-----TG- 417
Db 462 DAPSNLRFATTPNSLLVSWQPPRARITGYIIKYEKSPPREVVRPRPGVTEATITGL 521
QY 418 -PGQKAILFL-----PM--SAASDELPLQVTLPHNHLHGPEILDVPST 457
Db 522 EPGTEYTIIVIALKNNQKSEPLGRKKTDELPLQVTLPHNHLHGPEILDVPST 574

RESULT 15

ID ADD49015 standard; protein; 574 AA.
AC ADD49015;
XX
DT 15-JAN-2004 (first entry)
XX
DE Fibronectin fragment CH-296 #SEQ ID 12.
KW Immunomodulator; cytotoxic lymphocyte; fibronectin; CD8 positive cell;
KW adoptive immunotherapy.
XX Synthetic.
OS
XX WO2003080817-A1.
XX
XX PD 02-OCT-2003.
XX
XX PF 25-MAR-2003; 2003WO-JP003575.
XX
XX PR 25-MAR-2002; 2002JP-00084414.
XX
XX (TAKI) TAKARA BIO INC.
XX
XX PI Sagawa H, Ideno M, Kato I;
XX
XX WPI; 2003-865122/80.
XX
XX Method for producing cytotoxic lymphocytes, useful in adoptive
PT immunotherapy, comprises culturing in the presence of fibronectin or its
PT fragments.
XX
XX Claim 11; SEQ ID NO 12; 131pp; Japanese.

CC The invention relates to a method for producing cytotoxic lymphocytes
CC comprising induction, maintenance or dilution (dilution) and culturing
CC them in the presence of fibronectin and/or its fragments. Also disclosed
CC is a method for increasing the ratio of CD8 positive cells. The
CC fibronectin fragments are preferably optionally mutated sequences
CC comprising fully defined sequences of between 25 and 99 amino acids.
CC Cytotoxic lymphocytes of the invention are useful in adoptive
CC immunotherapy. In an example from the invention a 15 day old culture of
CC LAK cells was assayed to determine the ratio of CD8 positive cells. The

CC results showed that those cells subjected to fibronectin fragments and
CC anti CD3 antibody during the early and middle phase of induction had a
CC higher ratio of CD8 positive cells than in the control cells. The current
CC sequence represents a fibronectin fragment of the invention.

XX Sequence 574 AA;

Query Match 64.1%; Score 1536.5; DB 7; Length 574;
Best Local Similarity 58.0%; Pred. No. 3 5e-107;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PTDLFTNIGPDTMTVWAPPESIDLTNFLVRYSPVQNEEDVAELSIQSDNAVLTNLL 60
Db 1 PTDLFTNIGPDTMTVWAPPESIDLTNFLVRYSPVQNEEDVAELSIQSDNAVLTNLL 60
QY 61 PCTEVVSVSSVYQHESTPLGRQKTGLDSTGIDFSDITANSFTVHWIAPRAITGYR 120
Db 61 PCTEVVSVSSVYQHESTPLGRQKTGLDSTGIDFSDITANSFTVHWIAPRAITGYR 120
QY 121 IRHHPHFTSGRPREDRVPHSRNSITLTNLTGTEVVVSVIALNGREESPLLIQOQSTVSD 180
Db 121 IRHHPHFTSGRPREDRVPHSRNSITLTNLTGTEVVVSVIALNGREESPLLIQOQSTVSD 180
QY 181 VPRDLEVAATPTSLLSWDAVAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLLSWDAVAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTIIVYAVTGRGDSPASSKPIINRYTEIDKPSMAAGSITTL---PALPEDGSGA 297
Db 241 PGVDYTIIVYAVTGRGDSPASSKPIINRYTEIDKPSMAIPAPTDLKTQVTPTSLSAQW 300
QY 298 FPGGHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAEEGVVSIKGVCAN 357
Db 301 TPN-----VLTGYRVVTP-----KEKTPMKEINLAPDSSVVVSGLMVAT 344
QY 358 RY----LAWKEDGRLLASK-----CVTDEC----- 378
Db 345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETITISWRTKTETITG 401
QY 379 FFERLESNNYTYRS-----RKYT----- 398
Db 402 FOVDVAVPANGOTPIORTIKPDVRSYITGLQPGTDYKIYLYTLNDNARSPVVIDASTAI 461
QY 399 -----SWYVALKRTGOY-----KLGSK-----TG- 417
Db 462 DAPSNLRFATTPNSLLVSWQPPRARITGYIIKYEKSPPREVVRPRPGVTEATITGL 521
QY 418 -PGQKAILFL-----PM--SAASDELPLQVTLPHNHLHGPEILDVPST 457
Db 522 EPGTEYTIIVIALKNNQKSEPLGRKKTDELPLQVTLPHNHLHGPEILDVPST 574

Search completed: May 3, 2004, 13:09:44
Job time : 75.4946 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:07:19 ; Search time 23.4541 Seconds
(without alignments)
1005.924 Million cell updates/sec

Title: US-09-775-964-5
Perfect score: 2398
Sequence: 1 PDLRTNIGPDMRTWAP.....LVTLPHNHLGPEILDVPST 457

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2398	100.0	457	4	US-09-366-009-5
2	2398	100.0	457	4	US-08-809-156B-5
3	2257	94.1	432	1	US-07-959-369-8
4	2257	94.1	432	2	US-08-836-854-20
5	2257	94.1	432	4	US-09-366-009-4
6	2257	94.1	432	4	US-08-809-156B-4
7	2254	94.0	432	1	US-07-959-369-9
8	1536.5	64.1	574	2	US-08-836-854-21
9	1536.5	64.1	574	3	US-09-463-296-1
10	1536.5	64.1	574	4	US-09-366-009-24
11	1536.5	64.1	574	4	US-08-809-156B-24
12	1527	63.7	489	4	US-09-366-009-8
13	1527	63.7	489	4	US-08-809-156B-8
14	1499	62.5	2231	1	US-08-153-799-16
15	1485.5	61.9	302	2	US-08-836-854-5
16	1485.5	61.9	302	4	US-09-366-009-29
17	1485.5	61.9	302	4	US-08-809-156B-29
18	1474	61.5	2324	1	US-08-283-857-1
19	1474	61.5	2324	5	PCT-US95-09819-1
20	1474	61.5	2327	6	5455158-1
21	1474	61.5	2386	2	US-09-016-366A-12
22	1474	61.5	2446	2	US-08-551-356-2
23	1474	61.5	2446	5	PCT-US93-12687-2
24	1463.5	61.0	472	4	US-09-366-009-21
25	1463.5	61.0	472	4	US-08-809-156B-21
26	1463.5	61.0	549	1	US-08-836-854-11
27	1463.5	61.0	549	4	US-09-366-009-23

28	1463.5	61.0	549	4	US-08-809-156B-23	Sequence 23, Appl
29	1455	60.7	422	2	US-08-836-854-12	Sequence 12, Appl
30	1454	60.6	826	4	US-09-366-009-14	Sequence 14, Appl
31	1454	60.6	826	4	US-08-809-156B-14	Sequence 14, Appl
32	1453	60.6	446	2	US-08-836-854-15	Sequence 15, Appl
33	1450	60.5	332	2	US-08-836-854-13	Sequence 13, Appl
34	1448.5	60.4	464	2	US-08-836-854-19	Sequence 19, Appl
35	1448.5	60.4	464	4	US-09-366-009-7	Sequence 7, Appl
36	1448.5	60.4	464	4	US-08-809-156B-7	Sequence 9, Appl
37	1447.5	60.4	474	2	US-08-836-854-9	Sequence 9, Appl
38	1447	60.3	368	2	US-08-836-854-17	Sequence 17, Appl
39	1442	60.1	341	2	US-08-836-854-14	Sequence 14, Appl
40	1442	60.1	367	2	US-08-836-854-18	Sequence 18, Appl
41	1442	60.1	457	2	US-08-836-854-16	Sequence 22, Appl
42	1442	60.1	457	4	US-09-366-009-22	Sequence 22, Appl
43	1442	60.1	457	4	US-08-809-156B-22	Sequence 22, Appl
44	1437	59.9	277	1	US-07-959-369-3	Sequence 3, Appl
45	1437	59.9	279	1	US-07-959-369-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-366-009-5
; Sequence 5, Application US/09366009
; Patent No. 6426042
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6426042uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear

```
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-366-009-5

Query Match          100.0%; Score 2398; DB 4; Length 457;
Best Local Similarity 100.0%; Pred. No. 1.2e-171;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTDLRFNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLRFNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60

QY 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180

QY 181 VPRDLEVVAAATPTSLISWDAPAVTVYRYITVYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVVAAATPTSLISWDAPAVTVYRYITVYGETGNSPVQEFVPGSKSTATISGLK 240

QY 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFPP 300
DB 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFPP 300

QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGCANRYL 360

QY 361 AKMEDGRLASKCVTDCEFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420
DB 361 AKMEDGRLASKCVTDCEFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420

RESULT 2
US-08-809-156B-5
; Sequence 5, Application US/08090156B
; Patent No. 6472204
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; APPLICANT: Uemori, Takashi
; APPLICANT: Ueno, Takashi
; APPLICANT: Koyama, No. 6472204auto
; APPLICANT: Hashino, Kimikazu
; APPLICANT: Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; TITLE OF INVENTION: CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,156B
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03254
```

```
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-809-156B-5

Query Match          100.0%; Score 2398; DB 4; Length 457;
Best Local Similarity 100.0%; Pred. No. 1.2e-171;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTDLRFNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLRFNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60

QY 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180

QY 181 VPRDLEVVAAATPTSLISWDAPAVTVYRYITVYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVVAAATPTSLISWDAPAVTVYRYITVYGETGNSPVQEFVPGSKSTATISGLK 240

QY 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFPP 300
DB 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFPP 300

QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGCANRYL 360

QY 361 AKMEDGRLASKCVTDCEFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420
DB 361 AKMEDGRLASKCVTDCEFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420

RESULT 3
US-07-959-369-8
; Sequence 8, Application US/07959369
; Patent No. 5302701
; GENERAL INFORMATION:
; APPLICANT: Hidetaka HASHI et al.
; TITLE OF INVENTION: No. 5302701el Functional Polypeptide
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
```

;; COUNTRY: U.S.A.
;; ZIP: 20005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Wordperfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/959,369
;; FILING DATE: 19921013
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warren M. Cheek, Jr.
;; REGISTRATION NUMBER: 33,367
;; REFERENCE/DOCKET NUMBER:
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-371-8850
;; TELEFAX:
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 432 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: polypeptide
;; HYPOTHEetical:
;; ANTI-SENSE:
;; FRAGMENT TYPE:
;; ORIGINAL SOURCE:
;; ORGANISM:
;; STRAIN:
;; INDIVIDUAL ISOLATE:
;; DEVELOPMENTAL STAGE:
;; HAPLOTYPE:
;; TISSUE TYPE:
;; CELL TYPE:
;; CELL LINE:
;; ORGANELLE:
;; IMMEDIATE SOURCE:
;; LIBRARY:
;; CLONE:
;; POSITION IN GENOME:
;; CHROMOSOME/SEGMENT:
;; MAP POSITION:
;; UNITS:
;; FEATURE:
;; NAME/KEY:
;; LOCATION:
;; IDENTIFICATION METHOD:
;; OTHER INFORMATION:
;; PUBLICATION INFORMATION:
;; AUTHORS:
;; TITLE:
;; JOURNAL:
;; VOLUME:
;; ISSUE:
;; PAGES:
;; DATE:
;; DOCUMENT NUMBER:
;; FILING DATE:
;; PUBLICATION DATE:
;; RELEVANT RESIDUES IN SEQ ID NO:
US-07-959-369-8

Query Match 94.1%; Score 2257; DB 1; Length 432;
Best Local Similarity 99.8%; Pred. No. 3.9e-161;
Matches 431; Conservative 1; Indels 0; Gaps 0;

QY 1 PTDLRFTNIGPDMRVTWAPPPSIDLTNFLVRYSPVKNEEDVAELSISPSDNAVLTNLL 60

Db 1 PTDLRFTNIGPDMRVTWAPPPSIDLTNFLVRYSPVKNEEDVAELSISPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVVEQHESTPLRGRQKTGLDSDPTGIDFSDITANSFTVHVIAPRATITGYR 120
Db 61 PGTEYVSVSSVVEQHESTPLRGRQKTGLDSDPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTPTGTEYVSVISVALNGREESPLLICQOSTVSD 180
Db 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTPTGTEYVSVISVALNGREESPLLICQOSTVSD 180
QY 181 VPRDLEVVAATPTSLLSISWDAPAVTVYRYITVGETGNSPVQEFVPGSKSTATISGLK 240
Db 181 VPRDLEVVAATPTSLLSISWDAPAVTVYRYITVGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITIVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFPP 300
Db 241 PGVDYITIVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFPP 300
QY 301 GHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDPHIKLQQAERGVVSIKGVCANRYL 360
Db 301 GHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDPHIKLQQAERGVVSIKGVCANRYL 360
QY 361 ANKEDGRLLASKCVTDECFERLESNNYNTYRSRKYTSWYVALKETGYKLGSKTGPQG 420
Db 361 ANKEDGRLLASKCVTDECFERLESNNYNTYRSRKYTSWYVALKETGYKLGSKTGPQG 420
QY 421 KAILFLPMSAAS 432
Db 421 KAILFLPMSAKS 432

RESULT 4

US-08-836-854-20
; Sequence 20, Application US/08836854
; Patent No. 5824547
; GENERAL INFORMATION:
; APPLICANT: HASHINO, Kimikazu
; APPLICANT: MATSUSHITA, Hideyuki
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSPLECTED CELLS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W. Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,854
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP95/02425
; FILING DATE: 29-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 317721/1994
; FILING DATE: 29-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: HASHINO-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-836-854-20

Query Match          94.1%; Score 2257; DB 2; Length 432;
Best Local Similarity 99.8%; Pred. No. 3.9e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTDLRFNIGDPTMRVTWAPPSSIDLTNFLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
Db 1 PTDLRFNIGDPTMRVTWAPPSSIDLTNFLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
QY 61 PGTEYVVSVSSVYEQHESTPLRGQKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
Db 61 PGTEYVVSVSSVYEQHESTPLRGQKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
QY 121 IRHHPEHFGSRPREDRVPHSRNSITLTNLTGTEYVVSVI VALNGREESPLLIGQQSTVSD 180
Db 121 IRHHPEHFGSRPREDRVPHSRNSITLTNLTGTEYVVSVI VALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLIISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLIISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYITVYAVTGRGSDPASSKPI SINRYRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
Db 241 PGVDYITVYAVTGRGSDPASSKPI SINRYRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEBRGVVS IKGVCANRYL 360
Db 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEBRGVVS IKGVCANRYL 360
QY 361 AKMEDGRLLASKCVTDECFFERLESNNYTYRSRYTTSWYVALKRTGYKLGSKTGPQ 420
Db 361 AKMEDGRLLASKCVTDECFFERLESNNYTYRSRYTTSWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
Db 421 KAILFLPMSAKS 432
```

RESULT 5

```

US-09-366-009-4
; Sequence 4, Application US/09366009
; Patent No. 6426042
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6426042uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
```

```

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-366-009-4

Query Match          94.1%; Score 2257; DB 4; Length 432;
Best Local Similarity 99.8%; Pred. No. 3.9e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTDLRFNIGDPTMRVTWAPPSSIDLTNFLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
Db 1 PTDLRFNIGDPTMRVTWAPPSSIDLTNFLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
QY 61 PGTEYVVSVSSVYEQHESTPLRGQKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
Db 61 PGTEYVVSVSSVYEQHESTPLRGQKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
QY 121 IRHHPEHFGSRPREDRVPHSRNSITLTNLTGTEYVVSVI VALNGREESPLLIGQQSTVSD 180
Db 121 IRHHPEHFGSRPREDRVPHSRNSITLTNLTGTEYVVSVI VALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLIISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLIISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYITVYAVTGRGSDPASSKPI SINRYRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
Db 241 PGVDYITVYAVTGRGSDPASSKPI SINRYRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEBRGVVS IKGVCANRYL 360
Db 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEBRGVVS IKGVCANRYL 360
QY 361 AKMEDGRLLASKCVTDECFFERLESNNYTYRSRYTTSWYVALKRTGYKLGSKTGPQ 420
Db 361 AKMEDGRLLASKCVTDECFFERLESNNYTYRSRYTTSWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
Db 421 KAILFLPMSAKS 432
```

RESULT 6

```

US-08-809-156B-4
; Sequence 4, Application US/08809156B
; Patent No. 6472204
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6472204uto
; Hashino, Kimikazu
; Kato, Ikunoshin
```

;; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
;; TITLE OF INVENTION: CELLS WITH RETROVIRUS
;; NUMBER OF SEQUENCES: 39
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: WEISER & ASSOCIATES
;; STREET: 230 South Fifteenth Street, Suite 500
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19102
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/809,156B
;; FILING DATE: 07-MAR-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP96/03254
;; FILING DATE: 07-NOV-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 294382/1995
;; FILING DATE: 13-NOV-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 051847/1996
;; FILING DATE: 08-MAR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weiser, Gerard J.
;; REGISTRATION NUMBER: 19,763
;; REFERENCE/DOCKET NUMBER: 977,6507P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-875-8383
;; TELEFAX: 215-875-8394
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 432 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-08-809-156B-4

Query Match 94.1%; Score 2257; DB 4; Length 432;
Best Local Similarity 99.8%; Pred. No. 3.9e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PTDLFTNIGPDMRVMTWAPPIDLTNFLVRYSPVKNEEDVAELSISSPDNAVLTNLL 60
DB 1 PTDLFTNIGPDMRVMTWAPPIDLTNFLVRYSPVKNEEDVAELSISSPDNAVLTNLL 60
QY 61 PGTEYVSVSSVEQHESTPLRGOKTGLDSPGTIDFSDITANSFTVHMIAPRATITGYR 120
DB 61 PGTEYVSVSSVEQHESTPLRGOKTGLDSPGTIDFSDITANSFTVHMIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNITLTNTPGTEYVVSVIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNITLTNTPGTEYVVSVIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLVVAATPSTLLISWDAPAVTVRYRITYTGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLVVAATPSTLLISWDAPAVTVRYRITYTGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITTVAVTVGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFPP 300
DB 241 PGVDYITTVAVTVGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFPP 300
QY 301 GHFKDPKRLCYKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEGVVSIKGVCANRYL 360
DB 301 GHFKDPKRLCYKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEGVVSIKGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDECFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420

DB 361 AMKEDGRLLASKCVTDECFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
DB 421 KAILFLPMSAKS 432
RESULT 7
US-07-959-369-9
; Sequence 9, Application US/07959369
; Patent No. 5302701
; GENERAL INFORMATION:
; APPLICANT: Hidetaka HASHI et al.
; TITLE OF INVENTION: No. 5302701el Functional Polypeptide
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/959,369
; FILING DATE: 19921013
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:

IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:

Query Match 94.0%; Score 2254; DB 1; Length 432;
Best Local Similarity 99.5%; Pred. No. 6.6e-161;
Matches 430; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTDLFTNIGPDMRVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISGPSNNAVLTNLL 60
DB 1 PTDLFTNIGPDMRVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISGPSNNAVLTNLL 60
QY 61 PGTYYVSVSSVYEQHESTPLRGKQKTGLDSDPTGIDFSDITANSFTVHIAPRATITGYR 120
DB 61 PGTYYVSVSSVYEQHESTPLRGKQKTGLDSDPTGIDFSDITANSFTVHIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
QY 181 VPRDLEWVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEWVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
DB 241 PGVDYITVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
QY 301 GHFKDKRLYCKNGGFPLRIHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCANRYL 360
DB 301 GHFKDKRLYCKNGGFPLRIHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCANRYL 360
QY 361 AKKEDGRLASKCVTDCEFFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQ 420
DB 361 AKKEDGRLASKCVTDCEFFERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
DB 421 KAILFLPMSAKS 432

RESULT 8

US-08-836-854-21
Sequence 21, Application US/08836854
Patent No. 5824547
GENERAL INFORMATION:
APPLICANT: HASHINO, Kimikazu
APPLICANT: MATSUSHITA, Hideyuki
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W. Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836.854
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP95/02425
FILING DATE: 29-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 317721/1994
FILING DATE: 29-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: HASHINO=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-854-21

Query Match 64.1%; Score 1536.5; DB 2; Length 574;
Best Local Similarity 58.0%; Pred. No. 4.8e-107;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;
QY 1 PTDLFTNIGPDMRVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISGPSNNAVLTNLL 60
DB 1 PTDLFTNIGPDMRVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISGPSNNAVLTNLL 60
QY 61 PGTYYVSVSSVYEQHESTPLRGKQKTGLDSDPTGIDFSDITANSFTVHIAPRATITGYR 120
DB 61 PGTYYVSVSSVYEQHESTPLRGKQKTGLDSDPTGIDFSDITANSFTVHIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
QY 181 VPRDLEWVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEWVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTL---PALPEDGGSGA 297
DB 241 PGVDYITVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAIPAPTDLKTQVTPSLSAQW 300
QY 298 PPPGHFKDKRLYCKNGGFPLRIHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCAN 357
DB 301 TPPN-----VOLTGYRVVTP-----KEKTPMKKEINLAPDSSVVVSGLMVAT 344
QY 358 RY---LAKMEDGRLASK-----CVTDEC----- 378
DB 345 KYEVSVYALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETITISWRTKTETITG 401
QY 379 PFERLESNNYNTYRS-----RKYT----- 398
DB 402 FQDVAVPANGQPTIORTIKPDVRSYITGLQGTGYKYLYTLNDNARSSPVVIDASTAI 461
QY 399 -----SWYVALKRTGY-----KLASK-----TG- 417
DB 462 DAPSNLRFATTPNSLLVSWQPPRARITGYIIKYEKPGSPPREVVRPRPGVTEATITGL 521
QY 418 -PGQKAILFL-----PM--SAASDELPQVTLPHNPHLHGEIILDVPST 457
DB 522 EPGTEVTIVIALKNNQKSEPLIGRKTDELQVLTLPHPNHLHGEIILDVPST 574


```

RESULT 9
US-09-463-296-1
; Sequence 1, Application US/09463296
; Patent No. 6287864
; GENERAL INFORMATION:
; APPLICANT: BAGNIS, Claude
; APPLICANT: IMBERT, Anne-Marie
; APPLICANT: MANNONI, Patrice
; TITLE OF INVENTION: GENE TRANSFER METHOD WITH THE USE OF SERUM-FREE MEDIUM
; FILE REFERENCE: BAGNIS=1
; CURRENT APPLICATION NUMBER: US/09/463.296
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/JP98/03173
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: JP 196772/1997
; EARLIER FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-463-296-1

Query Match      64.1%; Score 1536.5; DB 3; Length 574;
Best Local Similarity 58.0%; Pred. No. 4.8e-107;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PDLRFTNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELSPSDNVAVLNLL 60
DB 1 PDLRFTNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELSPSDNVAVLNLL 60
QY 61 PGTEYVSVSSVYEQHESTPLRGKQKTGLDPSPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEQHESTPLRGKQKTGLDPSPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLICQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLICQSTVSD 180
QY 181 VPRDLEVAATPTSLLSWDAPAVTVRYRYITGTEGNSPVQEFVTPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLLSWDAPAVTVRYRYITGTEGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYTTIVYAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSITTL---PALPEDGSGA 297
DB 241 PGVDYTTIVYAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSITTL---PALPEDGSGA 297

US-09-366-009-24
; Sequence 1, Application US/09366009
; Patent No. 6426042
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; APPLICANT: Uemori, Takashi
; APPLICANT: Koyama, No. 6426042uto
; APPLICANT: Hashino, Kimikazu
; APPLICANT: Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; FILE REFERENCE: CELLS WITH RETROVIRUS
; CURRENT APPLICATION NUMBER: US/09/366.009
; CURRENT FILING DATE: 02-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-366-009-24

Query Match      64.1%; Score 1536.5; DB 4; Length 574;
Best Local Similarity 58.0%; Pred. No. 4.8e-107;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PDLRFTNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELSPSDNVAVLNLL 60
DB 1 PDLRFTNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELSPSDNVAVLNLL 60
QY 61 PGTEYVSVSSVYEQHESTPLRGKQKTGLDPSPTGIDFSDITANSFTVHVIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEQHESTPLRGKQKTGLDPSPTGIDFSDITANSFTVHVIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLICQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLICQSTVSD 180
QY 181 VPRDLEVAATPTSLLSWDAPAVTVRYRYITGTEGNSPVQEFVTPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLLSWDAPAVTVRYRYITGTEGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYTTIVYAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSITTL---PALPEDGSGA 297
DB 241 PGVDYTTIVYAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSITTL---PALPEDGSGA 297

```

RESULT 10

```

US-09-366-009-24
; Sequence 2, Application US/09366009
; Patent No. 6426042

```


;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; FILING APPLICATION NUMBER: US/09/366,009
;; FILING DATE: 02-Aug-1999
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/809,156
;; FILING DATE: <Unknown>
;; APPLICATION NUMBER: JP 294382/1995
;; FILING DATE: 13-NOV-1995
;; APPLICATION NUMBER: JP 051847/1996
;; FILING DATE: 08-MAR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weiser, Gerard J.
;; REGISTRATION NUMBER: 19,763
;; REFERENCE/DOCKET NUMBER: 977.6507P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-875-8383
;; TELEFAX: 215-875-8394
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 489 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: <Unknown>
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-366-009-8

Query Match 63.7%; Score 1527; DB 4; Length 489;
Best Local Similarity 65.5%; Pred. No. 2e-106;
Matches 342; Conservative 7; Mismatches 75; Indels 98; Gaps 11;
QY 1 PTDLRFNIGDPTMRVTWAPPSPDLTNFLVRYSPVKNEEDVAELSTSPSDNAVLTNLL 60
DB 1 PTDLRFNIGDPTMRVTWAPPSPDLTNFLVRYSPVKNEEDVAELSTSPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYEQHESTPLRGRQKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEQHESTPLRGRQKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVISVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVISVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVVAAATPTSLLSWDAPAVTVRYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
DB 181 VPRDLEVVAAATPTSLLSWDAPAVTVRYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYTTVYAVTGRGDSPASSKPIISINRYTEIDKPSMA-AGSITTLTLPALPEDGGSG- 296
DB 241 PGVDYTTVYAVTGRGDSPASSKPIISINRYTEIDKPSMGIRGLKGTGKEGDEGDFPGFKG 300
QY 297 -----APPPGHF-----KDPKRLYCKNG-----GFFLRIH 321
DB 301 DMGIKGRGEITGPGPRGEDGPEGPKRGPGNGDPGLPGPGEKGLGVLGPGLPYGRQG 360
QY 322 PGRVD-----GVREKSDPHIKLQLOAERGVSIVKVCANRYLAMKE--- 364
DB 361 PKGSIGFPFGANGEGKGRGTGPKGPR-----CQRGTPGRGERGRGITGPKGPK 413
QY 365 -----DGRLLASKCVTDCFFERLESNNYTSRKYTSWYVALKRTQYKLGSKTGP 419
DB 414 GNSGGDGPAGPPG-----ERGPNGPQG-----TG--FPGKGPFG 447
QY 420 QKAILFLP-----MSAASDELPLQVTLPHPLHGEILDPVST 457
DB 448 PPKDGLPCHPGQRGASDELPLQVTLPHPLHGEILDPVST 489

RESULT 13
US-08-809-156B-8

;; Sequence 8, Application US/08809156B
;; Patent No. 6472204
;; GENERAL INFORMATION:
;; APPLICANT: Asada, Kiyozo
;; APPLICANT: Umori, Takashi
;; APPLICANT: Ueno, Takashi
;; APPLICANT: Koyama, No. 6472204uto
;; APPLICANT: Hashino, Kimikazu
;; APPLICANT: Kato, Ikunoshin
;; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
;; TITLE OF INVENTION: CELLS WITH RETROVIRUS
;; NUMBER OF SEQUENCES: 39
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: WEISER & ASSOCIATES
;; STREET: 230 South Fifteenth Street, Suite 500
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19102
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/809,156B
;; FILING DATE: 07-MAR-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP96/03254
;; FILING DATE: 07-NOV-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 294382/1995
;; FILING DATE: 13-NOV-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 051847/1996
;; FILING DATE: 08-MAR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weiser, Gerard J.
;; REGISTRATION NUMBER: 19,763
;; REFERENCE/DOCKET NUMBER: 977.6507P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-875-8383
;; TELEFAX: 215-875-8394
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 489 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-809-156B-8

Query Match 63.7%; Score 1527; DB 4; Length 489;
Best Local Similarity 65.5%; Pred. No. 2e-106;
Matches 342; Conservative 7; Mismatches 75; Indels 98; Gaps 11;
QY 1 PTDLRFNIGDPTMRVTWAPPSPDLTNFLVRYSPVKNEEDVAELSTSPSDNAVLTNLL 60
DB 1 PTDLRFNIGDPTMRVTWAPPSPDLTNFLVRYSPVKNEEDVAELSTSPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYEQHESTPLRGRQKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEQHESTPLRGRQKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVISVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVISVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVVAAATPTSLLSWDAPAVTVRYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
DB 181 VPRDLEVVAAATPTSLLSWDAPAVTVRYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240

QY 241 PGVDYITVYAVTGRGDSPPASSKPISINRYTEIDKPSMA-AGSITTLPALPEDGGSG--- 296
Db 241 PGVDYITVYAVTGRGDSPPASSKPISINRYTEIDKPSMGIRGLKGTGKEGEGDPGPKG 300
QY 297 -----AFPPGHF-----KDPKLYCKNG-----GFFLRH 321
Db 301 DMGIKDRGIGPPRGEDGPKRGKGGPNPDGPLGPPGKGLGVPGPGYGRQG 360
QY 322 PDGRVD-----GVREKSDPHIKLOQAEERGVSIGKVCANRYLAMKE--- 364
Db 361 PKSGIGPPGPGANGKGGRTGPKCPR-----GQGPPTGRGERGRTGPKGPK 413
QY 365 -----DORLLASKVTDCEFFERLESNNVNTVRSRKYTSWYVALKRTGQYKLGSKTGP 419
Db 414 GNSGDDGAGPPG-----BRGPNPQGP-----TG--FPGKGP 447
QY 420 QKAILFLP-----NSAASDELQVLTLPHPNLHGPEILDVPST 457
Db 448 PPKDGLPGHPGORGASDELQVLTLPHPNLHGPEILDVPST 489

RESULT 14

US-08-153-799-16
; Sequence 16, Application US/08153799
; Patent No. 5766883
; GENERAL INFORMATION:
; APPLICANT: Ballance, David J
; APPLICANT: Goodey, Andrew R
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: R Hain Swope, BOC Health Care Inc
; STREET: 100 Mountain Avenue
; CITY: Murray Hill
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07974
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/153,799
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/847975
; FILING DATE: 06-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8909916.2
; FILING DATE: 29-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB90/00650
; FILING DATE: 26-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/775952
; FILING DATE: 29-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Swope, R Hain
; REGISTRATION NUMBER: 24864
; REFERENCE/DOCKET NUMBER: 92H832
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 665 2400
; TELEFAX: (908) 771 6159
; TELEX: 219484
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2231 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO

; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..2231
; OTHER INFORMATION: /note= "Human fibronectin"
US-08-153-799-16

Query Match 62.5%; Score 1499; DB 1; Length 2231;
Best Local Similarity 51.4%; Pred No. 1.8e-103;
Matches 346; Conservative 31; Mismatches 64; Indels 232; Gaps 17;

QY 1 PTDLFTNIGDPTMTVWAPPSIDLNTNPLVRYSPVKNEEDVAELISPSDNNAVLTNLL 60
Db 1239 PTDLFTNIGDPTMTVWAPPSIDLNTNPLVRYSPVKNEEDVAELISPSDNNAVLTNLL 1298

QY 61 PCTEYVVSVSVEQHESTPLRGROKGTGLDSTGTDFSDITANSTVTHIAPRAITGYR 120
Db 1299 PCTEYVVSVSVEQHESTPLRGROKGTGLDSTGTDFSDITANSTVTHIAPRAITGYR 1358

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGCTEYVVSIVALNGREESPLLIQOQSTVSD 180
Db 1359 IRHHPHFSGRPREDRVPHSRNSITLTNLTGCTEYVVSIVALNGREESPLLIQOQSTVSD 1418

QY 181 VPRDLEVVAAATPTSLLSWDAPAVTVRYRYITYGETGNSPVQVEFTVFGSKSTATISGLK 240
Db 1419 VPRDLEVVAAATPTSLLSWDAPAVTVRYRYITYGETGNSPVQVEFTVFGSKSTATISGLK 1478

QY 241 PGVDYITVYAVTGRGDSPPASSKPISINRYTEIDKPSMA-----AGSITTLPA--- 288
Db 1479 PGVDYITVYAVTGRGDSPPASSKPISINRYTEIDKPSMQVTDVQDNISVKNLPSSPV 1538

QY 289 -----LPEDGGSGAFPPGHFKDPK-----RLYCKN----- 313
Db 1539 TGYRVTTTPKNG-----PGTKTKTAGPDQTEMIEGLQPTVEYVVSIVAQNPGEQSP 1592

QY 314 -----GGFFLRHDPDGRVDGVREKSDPHI 337
Db 1593 LVQTAVTTPAPTDLKFQVTPTSLSAQWTPDVQLTGYRVVTP-----KQKTGPMK 1645

QY 338 KLOQAEERGVSIGKVCANRY-----LAMKEDGRLLASK-----CVT 375
Db 1646 EINLAPDSSSVVSGLMVATKYEVSVYALKD---TLTSRPAQGVVTTLENVSPRRARVT 1702

QY 376 DEC-----FFERLESNNYTVRS-----RKYT----- 398
Db 1703 DATETITISWRTKTETITGQVDVAVPANGQTPQRTIKPDVRSYITGLQGTDYKIYL 1762

QY 399 -----SWYVALKRTQY-----KLGSK 415
Db 1763 YTLNDNARSSPVVIDASTAIDAPSNLRLFLATTPNSLLVSWQPPRARITGYIKYKPGSP 1822

QY 416 -----TG--PGQKAILFL-----PM--SAASDELQVLTLPHP 444
Db 1823 PREVVRPRPGVTEATITGLEPFGTEYTIYVIALKNNKSEFLGRKTKDELQVLTLPHP 1882

QY 445 NLHGPEILDVPST 457
Db 1883 NLHGPEILDVPST 1895

RESULT 15

US-08-836-854-5
; Sequence 5, Application US/08836854
; Patent No. 5824547
; GENERAL INFORMATION:
; APPLICANT: HASHINO, Kimikazu
; APPLICANT: MATSUSHITA, Hideyuki
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W. Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,854
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP95/02425
FILING DATE: 29-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 317721/1994
FILING DATE: 29-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: HASHINO-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-854-5.

Query Match 61.9%; Score 1485.5; DB 2; Length 302;
Best Local Similarity 66.1%; Pred. No. 1.4e-103;
Matches 302; Conservative 0; Mismatches 0; Indels 155; Gaps 1;

Qy 1 PTDLRFTNIGPDTMRVTWAPPPSIDLTFNLVRYSPVKNEEDVAELISPSDNAVLTNLL 60
Db 1 PTDLRFTNIGPDTMRVTWAPPPSIDLTFNLVRYSPVKNEEDVAELISPSDNAVLTNLL 60

Qy 61 PGTEYVVSVSVEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120
Db 61 PGTEYVVSVSVEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120

Qy 121 IRHHPHFGRPREDRVPHSRNSITLNLTPGTEYVVSIVALNGREESPLLIQQSTVSD 180
Db 121 IRHHPHFGRPREDRVPHSRNSITLNLTPGTEYVVSIVALNGREESPLLIQQSTVSD 180

Qy 181 VPRDLEVAATPTSLLSWAPAVTVRYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLLSWAPAVTVRYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240

Qy 241 PGVDYTTIVAVTGRGSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGSGAFPP 300
Db 241 PGVDYTTIVAVTGRGSPASSKPIISINYTEIDK----- 276

Qy 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLOQAEERGVSIGVCANRYL 360
Db 277 ----- 276

Qy 361 AMKEDGRLALKCVTDCFFERLESNNYTYRSKTYSWYVALKRTGYKLGSKTGPQG 420
Db 277 ----- 276

Qy 421 KAILFLPMSAASDELPLQVTLPHNLHGPEILDVPST 457
Db 277 -----SDELPLQVTLPHNLHGPEILDVPST 302

Search completed: May 3, 2004, 13:12:34
Job time : 25.4541 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:11:14 ; Search time 54.7263 Seconds
(without alignments)
2314.713 Million cell updates/sec

Title: US-09-775-964-5
Perfect score: 2398
Sequence: 1 PDLRFTNIGPDMRTWAP.....LVTLPHPNLGHPILDVPST 457

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2398	100.0	457	10	US-09-775-964-5
2	2257	94.1	432	10	US-09-775-964-4
3	1536.5	64.1	574	10	US-09-775-964-24
4	1527	63.7	489	10	US-09-775-964-8
5	1485.5	61.9	302	10	US-09-775-964-29
6	1474	61.5	2320	12	US-10-236-392-2
7	1474	61.5	2320	14	US-10-279-733-8
8	1474	61.5	2328	12	US-10-182-936A-98
9	1474	61.5	2328	14	US-10-171-311-64
10	1474	61.5	2328	15	US-10-236-031B-70
11	1474	61.5	2328	15	US-10-374-979-98
12	1474	61.5	2355	15	US-10-144-194A-104
13	1474	61.5	2355	15	US-10-360-101-235
14	1474	61.5	2355	16	US-10-447-161-3
15	1474	61.5	2386	10	US-09-961-403-1

16	1463.5	61.0	472	10	US-09-775-964-21	Sequence 21, Appl
17	1463.5	61.0	549	10	US-09-775-964-23	Sequence 23, Appl
18	1454	60.6	826	10	US-09-775-964-14	Sequence 14, Appl
19	1448.5	60.4	464	10	US-09-775-964-7	Sequence 22, Appl
20	1442	60.1	457	10	US-09-775-964-22	Sequence 7, Appl
21	1437	59.9	599	14	US-10-279-733-9	Sequence 9, Appl
22	1437	59.9	2220	12	US-10-236-392-4	Sequence 4, Appl
23	1421	59.3	274	10	US-09-775-964-25	Sequence 25, Appl
24	840.5	35.1	501	9	US-09-934-706-4	Sequence 4, Appl
25	824.5	34.4	196	12	US-10-389-821-6	Sequence 6, Appl
26	824.5	34.4	210	9	US-09-902-773A-4	Sequence 4, Appl
27	824.5	34.4	210	12	US-10-372-653-8	Sequence 8, Appl
28	821	34.2	158	9	US-09-826-210-2	Sequence 2, Appl
29	820.5	34.2	386	12	US-10-344-634-14	Sequence 14, Appl
30	820	34.2	155	9	US-09-822-485-5	Sequence 5, Appl
31	820	34.2	155	9	US-09-802-365-8	Sequence 8, Appl
32	820	34.2	155	9	US-09-251-263-10	Sequence 10, Appl
33	820	34.2	155	9	US-09-425-021-10	Sequence 10, Appl
34	820	34.2	155	9	US-09-886-856-8	Sequence 8, Appl
35	820	34.2	155	9	US-09-749-728B-7	Sequence 7, Appl
36	820	34.2	155	10	US-09-902-460-2	Sequence 2, Appl
37	820	34.2	155	10	US-09-345-373-17	Sequence 17, Appl
38	820	34.2	155	10	US-09-775-964-3	Sequence 3, Appl
39	820	34.2	155	14	US-10-081-347-28	Sequence 28, Appl
40	820	34.2	155	14	US-10-075-446-17	Sequence 17, Appl
41	820	34.2	155	14	US-10-189-360-11	Sequence 11, Appl
42	820	34.2	155	14	US-10-192-988-14	Sequence 14, Appl
43	820	34.2	155	14	US-10-374-207-5	Sequence 5, Appl
44	820	34.2	155	14	US-10-395-541-5	Sequence 5, Appl
45	820	34.2	155	14	US-10-123-481-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-775-964-5
; Sequence 5, Application US/09775964
; Publication No. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Ueno, Takashi
; Koyama, No. US20030087437A1uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/775,964
FILING DATE: 20-Feb-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/366,009
FILING DATE: 02-Aug-1999
APPLICATION NUMBER: 08/809,156
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 294382/1995
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: JP 051847/1996

RESULT 3

US-09-775-964-24

; Sequence 24, Application US/09775964

; Publication No. US20030087437A1

; GENERAL INFORMATION:

; APPLICANT: Asada, Kiyozo

; Uemori, Takashi

; Koyama, No. US20030087437A1uto

; Hashino, Kimikazu

; Kato, Ikunoshin

; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET

; CELLS WITH RETROVIRUS

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: WEISER & ASSOCIATES

; STREET: 230 South Fifteenth Street, Suite 500

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/775,964

; FILING DATE: 20-Feb-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/366,009

; FILING DATE: 02-Aug-1999

; APPLICATION NUMBER: 08/809,156

; FILING DATE: <Unknown>

; APPLICATION NUMBER: JP 294382/1995

; FILING DATE: 13-NOV-1995

; APPLICATION NUMBER: JP 051847/1996

; FILING DATE: 08-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Weiser, Gerard J.

; REGISTRATION NUMBER: 19,763

; REFERENCE/DOCKET NUMBER: 977.6507P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-875-8383

; TELEFAX: 215-875-8394

; INFORMATION FOR SEQ ID NO: 24:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 574 amino acids

; TYPE: amino acid

; STRANDEDNESS: <Unknown>

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 24:

US-09-775-964-24

Query Match 64.1%; Score 1536.5; DB 10; Length 574;
Best Local Similarity 58.0%; Pred. No. 2e-115;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PTDLRFTNIGDTRVTVWAPPSPSIDLTNFLVRYSPVKNEEDVAELSIKPSDNAVLTNLL 60
DB 1 PTDLRFTNIGDTRVTVWAPPSPSIDLTNFLVRYSPVKNEEDVAELSIKPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYQHESTPLRQKTKGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
DB 61 PGTEYVSVSSVYQHESTPLRQKTKGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
QY 121 IRHHPEHFGSRPREDRVPHSRNITLTNLTGTEYVSVISVALNGREESPLLICQQTVD 180
DB 121 IRHHPEHFGSRPREDRVPHSRNITLTNLTGTEYVSVISVALNGREESPLLICQQTVD 180

QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYTIIVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTL---PALPEDGSSGA 297
DB 241 PGVDYTIIVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAIPAPFTDLKFTQVTPPTSLSAQW 300
QY 298 FPPGHFKDPKRLYCKNGGFFLRHDPGRVDGVRKSDPHIKLQAEERGVWSIKGVCAN 357
DB 301 TPN-----VQLTGYRVVTP-----KEKTPMKKEINLADSSSVVVSGLMVAT 344
QY 358 RY-----LAKKEDGRLLASK-----CVTDEC----- 378
DB 345 KYEVSVYALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETITISWRTKTETITG 401
QY 379 FFERLESNNYTYRS-----RKYT----- 398
DB 402 FQVDAVPANGQTPIQRTIKPDRVSYYTITGLQPGTDYKIYLYTLNDNARSSPVVIDASTAI 461
QY 399 -----SMYVALKRTGOV-----KLGSK-----TG- 417
DB 462 DAPSNURFLATTPNSLLVSWQPPRARITGVIIKYEKPGSPPREVVRPRRCVTEATITGL 521
QY 418 -PQKAILFL-----PM---SAASDELQVLTLPHPNLHGPEILDVPST 457
DB 522 EPQTEVTIVIALKNQKSEPLIGRKKTDLPQVLTLPHPNLHGPEILDVPST 574

RESULT 4
US-09-775-964-8
; Sequence 8, Application US/09775964
; Publication No. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. US20030087437A1uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/775,964
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-775-964-8

Query Match 63.7%; Score 1527; DB 10; Length 489;
Best Local Similarity 65.5%; Pred. No. 9.5e-115;
Matches 342; Conservative 7; Mismatches 75; Indels 98; Gaps 11;
QY 1 PTDLRFNIGDPTMRVTWAPPSSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLRFNIGDPTMRVTWAPPSSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYEOHESPTLRGQKTGLDPTGIDFSDITANSFTVHIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEOHESPTLRGQKTGLDPTGIDFSDITANSFTVHIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTIITVAVTGRGSDPASSKPIISINRYTEIDKPSMA-AGSITLTPALPEDGGSG--- 296
DB 241 PGVDYTIITVAVTGRGSDPASSKPIISINRYTEIDKPSMA-AGSITLTPALPEDGGSG--- 296
QY 297 -----AFPPGHF-----KDPKRLCYKNG-----GFFLRH 321
DB 301 DMGIKGRGIBGIPGPRGEDGPGKGRGPGNDGPGPLGPGKGLGVPGLPGYFGROG 360
QY 322 PDGRVD-----GVREKSDPHIKLOQAEERGVVISIKVCANRYLAMKE----- 364
DB 361 PKGSI GPFPGANGKGGRTGPKGPR-----GQRGPTGPRGERGPRGITGKPGPK 413
QY 365 -----DGRLLASCVTDECFEERLESNNYNTYRSRKYTSWYVALKRTQYKLGSKTGP 419
DB 414 GNSGGDGPAGPPG-----ERGNPGQGP-----TG--FPGKGP 447
QY 420 QKAILFLP-----MSAASDELQVLVTLPHNHLHGPEILDVPST 457
DB 448 PPGKDGLPHPGQRGASDELQVLVTLPHNHLHGPEILDVPST 489

RESULT 5
US-09-775-964-29
Sequence 29, Application US/09775964
Publication No. US20030087437A1
GENERAL INFORMATION:
APPLICANT: Asada, Kiyozo
Umori, Takashi
Ueno, Takashi
Koyama, No. US20030087437A1
Hashino, Kimikazu
Kato, Ikunoshin
TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
CELLS WITH RETROVIRUS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: WEISER & ASSOCIATES
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: PA

COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/775,964
FILING DATE: 20-Feb-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/366,009
FILING DATE: 02-Aug-1999
APPLICATION NUMBER: 08/809,156
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 294382/1995
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: JP 051847/1996
FILING DATE: 08-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 977.6507P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-09-775-964-29

Query Match 61.9%; Score 1485.5; DB 10; Length 302;
Best Local Similarity 66.1%; Pred. No. 1.1e-111;
Matches 302; Conservative 0; Mismatches 0; Indels 155; Gaps 1;
QY 1 PTDLFTNIGDPTMRVTWAPPSSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLFTNIGDPTMRVTWAPPSSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYEOHESPTLRGQKTGLDPTGIDFSDITANSFTVHIAPRATITGYR 120
DB 61 PGTEYVSVSSVYEOHESPTLRGQKTGLDPTGIDFSDITANSFTVHIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTIITVAVTGRGSDPASSKPIISINRYTEIDKPSMAAGSIITLTPALPEDGGSGAFP 300
DB 241 PGVDYTIITVAVTGRGSDPASSKPIISINRYTEIDKPSMAAGSIITLTPALPEDGGSGAFP 300
QY 301 GHFKDPKRLCYKNGGFFLRHDPGRVDGVRKSDPHIKLOQAEERGVVISIKVCANRYL 360
DB 277 -----SDELQVLVTLPHNHLHGPEILDVPST 457
QY 421 KAILFLPMSAASDELQVLVTLPHNHLHGPEILDVPST 457
DB 277 -----SDELQVLVTLPHNHLHGPEILDVPST 457

US-10-279-733-8

```
Query Match      61.5%; Score 1474; DB 14; Length 2320;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY 1 PTDLRFTNIGPDTMRVTWAPPSSIDLTLNFLVRYSPVKNEEDVAELSISSDNVAVLTNLL 60
Db 1235 PTDLRFTNIGPDTMRVTWAPPSSIDLTLNFLVRYSPVKNEEDVAELSISSDNVAVLTNLL 1294

QY 61 PGTETVSVSSVYEQHESHTPLGRQKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 120
Db 1295 PGTETVSVSSVYEQHESHTPLGRQKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 1354

QY 121 IRHHPHFSGRPREDVRPHSRNSITLTLNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 180
Db 1355 IRHHPHFSGRPREDVRPHSRNSITLTLNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 1414

QY 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
Db 1415 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 1474

QY 241 PGVDYTIIVYAVTGRGDSPASSKPIISINRYTEIDKPS -----M 278
Db 1475 PGVDYTIIVYAVTGRGDSPASSKPIISINRYTEIDKPSQMQVTDVQDINSVKWLPSSGPV 1534

QY 279 AAGSITTLTP----- 287
Db 1535 TGVRVTTTPKNGPGPKTKTAGPDQTEMTIEGLQFTVEYVSVYAQNPSGESQPLVQTAV 1594

QY 288 -----ALPEDGGSGAPP-PGHFKDPKRL 309
Db 1595 TNIDRPKGLAFTDVDVDSIKIAWESPOQVSRVRYTYSPPEDGIHELFPAPDGEEDTAEL 1654

QY 310 YCKNGG-----FFLRH-----PDGRV 326
Db 1655 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIAPTDLKFTQVTPTSLSAQWTPPNVQL 1714

QY 327 DGVR-----EKSDPHIKLQQAERGVSIVKVCANRY-----LAMEDGRLLASK- 372
Db 1715 TGVRVRVTPKEKTGPMKEINLAPDSSVVVSGLMVATKYEVSVYALKD---TLTSRPAQG 1771

QY 373 -----CVTDEC-----PFFERLESNNYTVRS-----R 395
Db 1772 VVTTLNVSPRRARVTDATETTTISWRTKTETITGFOVDAPVANGQTPQRTIKPDVR 1831

QY 396 KYT-----SWYVA 403
Db 1832 SYTITGLQPGTDYKIYLTLDNARSSPVVIDASTAIDAPSNLRLFLATTPNSLLVSWQPP 1891

QY 404 LKRTGQY-----KLGSK-----TG--PGOKAILFL-----PM- 428
Db 1892 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLI 1951

QY 429 -SAASDELQPLVTLPHNHLHGPEILDVPST 457
Db 1952 GRKKTDELQPLVTLPHNHLHGPEILDVPST 1981
```

RESULT 8

```
US-10-182-936A-98
; Sequence 98, Application US/10182936A
; Publication NO. US20040038860A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Kristina M.
; APPLICANT: Anisowicz, Anthony
; APPLICANT: Bhat, Sheem
; APPLICANT: Damagnez, Veronique
; APPLICANT: Robinson, John
; APPLICANT: Yavorsky, Paul
; TITLE OF INVENTION: Reagents and Method for Modulating DKK-Mediated Interactions
; FILE REFERENCE: 032796-143
; CURRENT APPLICATION NUMBER: US/10/182,936A
```

```
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: PCT/US02/15982
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: US 60/291,311
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/353,058
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: US 60/361,293
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 2328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-182-936A-98

Query Match      61.5%; Score 1474; DB 12; Length 2328;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY 1 PTDLRFTNIGPDTMRVTWAPPSSIDLTLNFLVRYSPVKNEEDVAELSISSDNVAVLTNLL 60
Db 1243 PTDLRFTNIGPDTMRVTWAPPSSIDLTLNFLVRYSPVKNEEDVAELSISSDNVAVLTNLL 1302

QY 61 PGTETVSVSSVYEQHESHTPLGRQKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 120
Db 1303 PGTETVSVSSVYEQHESHTPLGRQKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 1362

QY 121 IRHHPHFSGRPREDVRPHSRNSITLTLNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 180
Db 1363 IRHHPHFSGRPREDVRPHSRNSITLTLNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 1422

QY 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
Db 1423 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 1482

QY 241 PGVDYTIIVYAVTGRGDSPASSKPIISINRYTEIDKPS -----M 278
Db 1483 PGVDYTIIVYAVTGRGDSPASSKPIISINRYTEIDKPSQMQVTDVQDINSVKWLPSSSPV 1542

QY 279 AAGSITTLTP----- 287
Db 1543 TGVRVTTTPKNGPGPKTKTAGPDQTEMTIEGLQFTVEYVSVYAQNPSGESQPLVQTAV 1602

QY 288 -----ALPEDGGSGAPP-PGHFKDPKRL 309
Db 1603 TNIDRPKGLAFTDVDVDSIKIAWESPOQVSRVRYTYSPPEDGIHELFPAPDGEEDTAEL 1662

QY 310 YCKNGG-----FFLRH-----PDGRV 326
Db 1663 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIAPTDLKFTQVTPTSLSAQWTPPNVQL 1722

QY 327 DGVR-----EKSDPHIKLQQAERGVSIVKVCANRY-----LAMEDGRLLASK- 372
Db 1723 TGVRVRVTPKEKTGPMKEINLAPDSSVVVSGLMVATKYEVSVYALKD---TLTSRPAQG 1779

QY 373 -----CVTDEC-----PFFERLESNNYTVRS-----R 395
Db 1780 VVTTLNVSPRRARVTDATETTTISWRTKTETITGFOVDAPVANGQTPQRTIKPDVR 1839

QY 396 KYT-----SWYVA 403
Db 1840 SYTITGLQPGTDYKIYLTLDNARSSPVVIDASTAIDAPSNLRLFLATTPNSLLVSWQPP 1899

QY 404 LKRTGQY-----KLGSK-----TG--PGOKAILFL-----PM- 428
Db 1900 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLI 1959

QY 429 -SAASDELQPLVTLPHNHLHGPEILDVPST 457
Db 1960 GRKKTDELQPLVTLPHNHLHGPEILDVPST 1989
```

```
RESULT 9
US-10-171-311-64
; Sequence 64, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerhsh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 2328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-64

Query Match 61.5%; Score 1474; DB 14; Length 2328;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY 1 PTDLFTNIGDPTMRVTWAPPSSIDLTLNPLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
DB 1243 PTDLFTNIGDPTMRVTWAPPSSIDLTLNPLVRYSPVKNEEDVAELSISSDNAVLTNLL 1302

QY 61 PGTEYVSVSSVYEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHIAPRATITGYR 120
DB 1303 PGTEYVSVSSVYEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHIAPRATITGYR 1362

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 1363 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 1422

QY 181 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 1423 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 1482

QY 241 PGVDYITVYAVTGRGDSPASSKPSINRYTEIDKPS-----M 278
DB 1483 PGVDYITVYAVTGRGDSPASSKPSINRYTEIDKPSQMVQTDVQDNSISVKWLPSSSPV 1542

QY 279 AAGSITTL----- 287
DB 1543 TGYRVTTTPKNGPGPKTKTAGDQTEMTEIGLOPTVEYVSVYAQNPGSGESQPLVQTAV 1602

QY 288 ----- 287
DB 1603 TNIDRPGKLAFTDVIDSIKAMESPQQGVSRVRYTSSPDGSIHELFPAPDGEETAEL 1662

QY 310 YCKNGG-----PFLRIH----- 326
DB 1663 QGLRPGSEYTVSVVALHDDMESOPLIGTOSTAIPATDLKFTQVTPSLSAQWTPENVOL 1722

QY 327 DGYR-----EKSDPHIKLOLAERGVSIVKVCANRY-----LAKEDGRLLASK----- 372
DB 1723 TGYRVRTPKBKTGPMKEINLAPDSSSVVSGMLMVAIKVEVSVALKD---TLTSRPAQ 1779
```

```
QY 373 -----CVTDEC-----FFFERLESNNYNTYS-----R 395
DB 1780 VVTTLENVSPRRARVTDATEITITISWRTKTETITGTFQVDVAPANGQTPQRTIKPDVR 1839

QY 396 KYT-----SWYVA 403
DB 1840 SYTITGLQPGTDYKIYLYTLNDNARSSPVVIDASTAIDAPSNRFLATTNSLLVSNQPP 1899

QY 404 LKRTGQY-----KLASK-----TG--PGQKAILFL-----PM- 428
DB 1900 RARITGYIIKYKPGSPPREVVPRPPRGVTEATITGLEGTETIIVIALKNNQKSEPLI 1959

QY 429 -SAASDELQVLTLPHNHLHGPEILDVPST 457
DB 1960 GRKKTDELQVLTLPHNHLHGPEILDVPST 1989

RESULT 10
US-10-236-031B-70
; Sequence 70, Application US/10236031B
; Publication No. US20030219760A1
; GENERAL INFORMATION:
; APPLICANT: Gordon, Gavin J.
; APPLICANT: Jensen, Roderick V.
; APPLICANT: Gullans, Steven R.
; APPLICANT: Bueno, Raphael
; TITLE OF INVENTION: Diagnostic and Prognostic Tests
; FILE REFERENCE: B00801/70265 (JRV/JAV)
; CURRENT APPLICATION NUMBER: US/10/236,031B
; CURRENT FILING DATE: 2002-09-05
; PRIOR APPLICATION NUMBER: US 60/317,389
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/407,431
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 2328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-031B-70

Query Match 61.5%; Score 1474; DB 15; Length 2328;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY 1 PTDLFTNIGDPTMRVTWAPPSSIDLTLNPLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
DB 1243 PTDLFTNIGDPTMRVTWAPPSSIDLTLNPLVRYSPVKNEEDVAELSISSDNAVLTNLL 1302

QY 61 PGTEYVSVSSVYEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHIAPRATITGYR 120
DB 1303 PGTEYVSVSSVYEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHIAPRATITGYR 1362

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 1363 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 1422

QY 181 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 1423 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 1482

QY 241 PGVDYITVYAVTGRGDSPASSKPSINRYTEIDKPS-----M 278
DB 1483 PGVDYITVYAVTGRGDSPASSKPSINRYTEIDKPSQMVQTDVQDNSISVKWLPSSSPV 1542

QY 279 AAGSITTL----- 287
DB 1543 TGYRVTTTPKNGPGPKTKTAGDQTEMTEIGLOPTVEYVSVYAQNPGSGESQPLVQTAV 1602

QY 288 ----- 287
DB 1603 TNIDRPGKLAFTDVIDSIKAMESPQQGVSRVRYTSSPDGSIHELFPAPDGEETAEL 1662

QY 310 YCKNGG-----PFLRIH----- 326
DB 1663 QGLRPGSEYTVSVVALHDDMESOPLIGTOSTAIPATDLKFTQVTPSLSAQWTPENVOL 1722

QY 327 DGYR-----EKSDPHIKLOLAERGVSIVKVCANRY-----LAKEDGRLLASK----- 372
DB 1723 TGYRVRTPKBKTGPMKEINLAPDSSSVVSGMLMVAIKVEVSVALKD---TLTSRPAQ 1779
```

Db 1603 TNIDRPKGLAFTDVDVDSIKIAWESPOGQVSRVRYTSSPEDGIHELFPAPDGEDTAEL 1662
Qy 310 YKNGG-----FFLRH-----PGRV 326
Db 1663 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIPTAPDGLKFTQVTPSLSAQWTPPNVOL 1722
Qy 327 DGVR-----EKSDPHIKLQQAEBRGVSIKGVCANRY-----LAKMEDGRLLASK----- 372
Db 1723 TGYRVVTPKKTGPMKEINLAPDSSVVVSLMWATKYEVSVYALKD---TLTSRPAQG 1779
Qy 373 -----CVTDEC-----FFERLESNNNTYRS-----R 395
Db 1780 VVTTLENVSPRRARVTDATETITISWRTKTETITGQVDAVPANGQTPQRTIKPDVR 1839
Qy 396 KYT-----SWVA 403
Db 1840 SYTITGLQPGTDYKIYLYTLNDNARSPVVIDASTAIDAPSNLRFATTPNSLLVSWQPP 1899
Qy 404 LKRTGQY-----KLGSK-----TG--PGKAILFL-----PM- 428
Db 1900 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNOKSEPLI 1959
Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1960 GRKKTDELQVLTLPHPNLHGPEILDVPST 1989

RESULT 11
US-10-374-979-98
; Sequence 98, Application US/10374979
; Publication No. US20030219793A1
; GENERAL INFORMATION:
; APPLICANT: John P. Carulli et al.
; TITLE OF INVENTION: THE HIGH BONE MASS GENE OF 11q13.3
; FILE REFERENCE: 032796-021
; CURRENT APPLICATION NUMBER: US/10/374,979
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 09/544,398
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 09/543,771
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 09/229,319
; PRIOR FILING DATE: 1999-01-13
; PRIOR APPLICATION NUMBER: US 60/071,449
; PRIOR FILING DATE: 1998-01-13
; PRIOR APPLICATION NUMBER: US 60/105,511
; PRIOR FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 109
; SEQ ID NO 98
; LENGTH: 2328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-374-979-98

Query Match 61.5%; Score 1474; DB 15; Length 2328;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLRFTNIGDPTMRVTWAPPSPIDLTNFLVRYSPVKNEEDVAELSISSPDNAVLTNLL 60
Db 1243 PTDLRFTNIGDPTMRVTWAPPSPIDLTNFLVRYSPVKNEEDVAELSISSPDNAVLTNLL 1302
Qy 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDSPGIDFSDITANSFTVHWIAPRATITGYR 120
Db 1303 PGTEYVSVSSVVEQHESTPLRGQKTGLDSPGIDFSDITANSFTVHWIAPRATITGYR 1362
Qy 121 IRHHPHFSGRPREDRVPHSRNITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180
Db 1363 IRHHPHFSGRPREDRVPHSRNITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 1422
Qy 181 VPRDLEVAATPTSLLSWDAPAVTVRYRITYTGTGGNSPVQEFVPGSKSTATISGLK 240
Db 1423 VPRDLEVAATPTSLLSWDAPAVTVRYRITYTGTGGNSPVQEFVPGSKSTATISGLK 1482

Qy 241 PGVDYTTTVAVTCRGDSPASSKPIISINYTEIDKPS-----M 278
Db 1483 PGVDYTTTVAVTCRGDSPASSKPIISINYTEIDKPSQMVTDQDINSISVKWLPSSSPV 1542
Qy 279 AAGSITTLP----- 287
Db 1543 TGYRVVTPKNGPQTKTAGPDQTEMTIEGLQPTVEYVSVVAQNPSGESQPLVQTAV 1602
Qy 288 -----ALPEDGGSGAFP--PGHFKDPKRL 309
Db 1603 TNIDRPKGLAFTDVDVDSIKIAWESPOGQVSRVRYTSSPEDGIHELFPAPDGEDTAEL 1662
Qy 310 YKNGG-----FFLRH-----PGRV 326
Db 1663 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIPTAPDGLKFTQVTPSLSAQWTPPNVOL 1722
Qy 327 DGVR-----EKSDPHIKLQQAEBRGVSIKGVCANRY-----LAKMEDGRLLASK----- 372
Db 1723 TGYRVVTPKKTGPMKEINLAPDSSVVVSLMWATKYEVSVYALKD---TLTSRPAQG 1779
Qy 373 -----CVTDEC-----FFERLESNNNTYRS-----R 395
Db 1780 VVTTLENVSPRRARVTDATETITISWRTKTETITGQVDAVPANGQTPQRTIKPDVR 1839
Qy 396 KYT-----SWVA 403
Db 1840 SYTITGLQPGTDYKIYLYTLNDNARSPVVIDASTAIDAPSNLRFATTPNSLLVSWQPP 1899
Qy 404 LKRTGQY-----KLGSK-----TG--PGKAILFL-----PM- 428
Db 1900 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNOKSEPLI 1959
Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1960 GRKKTDELQVLTLPHPNLHGPEILDVPST 1989

RESULT 12

US-10-144-194A-104
; Sequence 104, Application US/10144194A
; Publication No. US20030215809A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Breast Cancer Genes
; FILE REFERENCE: 3U 103 R1
; CURRENT APPLICATION NUMBER: US/10/144,194A
; CURRENT FILING DATE: 2002-06-12
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 104
; LENGTH: 2355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-194A-104

Query Match 61.5%; Score 1474; DB 15; Length 2355;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLRFTNIGDPTMRVTWAPPSPIDLTNFLVRYSPVKNEEDVAELSISSPDNAVLTNLL 60
Db 1270 PTDLRFTNIGDPTMRVTWAPPSPIDLTNFLVRYSPVKNEEDVAELSISSPDNAVLTNLL 1329
Qy 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDSPGIDFSDITANSFTVHWIAPRATITGYR 120
Db 1330 PGTEYVSVSSVVEQHESTPLRGQKTGLDSPGIDFSDITANSFTVHWIAPRATITGYR 1389
Qy 121 IRHHPHFSGRPREDRVPHSRNITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180
Db 1390 IRHHPHFSGRPREDRVPHSRNITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 1449
Qy 181 VPRDLEVAATPTSLLSWDAPAVTVRYRITYTGTGGNSPVQEFVPGSKSTATISGLK 240

Db 1450 VPRLEVAATPTSLISWDAPAVTVYRITYTGETGNSPVQETVPGSKSTATISGLK 1509
QY 241 PGVDYTTTAVTGRGDSPASSKPIISINYTEIDKPS-----M 278
Db 1510 PGVDYTTTAVTGRGDSPASSKPIISINYTEIDKPSQMVQVDVQDNSISVKWLPSSSPV 1569
QY 279 AAGSITLTP----- 287
Db 1570 TGRVVTTPKNGPGPTKKTAGPDQTEMTEIQLQPTVEYVSVVAQNPGSGESQPLVQTAV 1629
QY 288 -----ALPEDGSGGAFPP-PGHFKDPKRL 309
Db 1630 TNIDRPKGLAFTDVDVDSIKIAMESPQGVSRVRYTVSSPEDGIHELFPADGGEEDTAEL 1689
QY 310 YCKNGG-----FFLRIH-----PDGRV 326
Db 1690 QGLRPGSEYTVSVVVALHDDMESQPLIGTQSTAIPTDPAKFTQVTPPTSLSAQWTPPNVQL 1749
QY 327 DGVR-----EKSDPHIKLOQAEERGVSIGVCANRY-----LANKEDGRLLASK----- 372
Db 1750 TGRVVRVTCKEKTGPMKEINLAPDSSSVVSGMLVATKYEVSVVALKD---TLTSRPAQ 1806
QY 373 -----CVTDEC-----FFERLESNNYNTYRS-----R 395
Db 1807 VVTTLENVSPRRARVTDATETITISWRKTETITIGFQVDVAVPANGQTPQRTIKPDVR 1866
QY 396 KYT-----SWYVA 403
Db 1867 SYITITGLQPGTDYKIYLYTLNDNARSPPVVIDASTAIDAPSNLRFLLATTPNSLLVSWQPP 1926
QY 404 LKRTGQY-----KLGSK-----TG--PGQKAILFL-----PM- 428
Db 1927 RARITGYIIKYEKPGSPRPVPRPGVTEATITGLEPGTEYTIYVIALKNNKSEPLI 1986
QY 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1987 GRKKTDELQVLTLPHPNLHGPEILDVPST 2016

RESULT 13

US-10-360-101-235
; Sequence 235, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 235
; LENGTH: 2355
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of fibronectin
US-10-360-101-235

Query Match 61.5%; Score 1474; DB 15; Length 2355;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;
QY 1 PTDLRFNIGPDMRTVMTWAPPSPSIDLTNFLVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60
Db 1239 PTDLRFNIGPDMRTVMTWAPPSPSIDLTNFLVRYSPVKNEEDVAELSISSPSDNAVLTNLL 1298
QY 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDPSGTIDFSDITANSFTVHVIAPRATITGYR 120

Db 1299 PGTEYVSVSSVVEQHESTPLRGQKTGLDPSGTIDFSDITANSFTVHVIAPRATITGYR 1358
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTPTGTEYVSVI VALNGREESPLLIGQOSTVSD 180
Db 1359 IRHHPHFSGRPREDRVPHSRNSITLTNLTPTGTEYVSVI VALNGREESPLLIGQOSTVSD 1418
QY 181 VPRDLEVAATPTSLISWDAPAVTVYRITYTGETGNSPVQETVPGSKSTATISGLK 240
Db 1419 VPRDLEVAATPTSLISWDAPAVTVYRITYTGETGNSPVQETVPGSKSTATISGLK 1478
QY 241 PGVDYTTTAVTGRGDSPASSKPIISINYTEIDKPS-----M 278
Db 1479 PGVDYTTTAVTGRGDSPASSKPIISINYTEIDKPSQMVQVDVQDNSISVKWLPSSSPV 1538
QY 279 AAGSITLTP----- 287
Db 1539 TGRVVTTPKNGPGPTKKTAGPDQTEMTEIQLQPTVEYVSVVAQNPGSGESQPLVQTAV 1598
QY 288 -----ALPEDGSGGAFPP-PGHFKDPKRL 309
Db 1599 TNIDRPKGLAFTDVDVDSIKIAMESPQGVSRVRYTVSSPEDGIHELFPADGGEEDTAEL 1658
QY 310 YCKNGG-----FFLRIH-----PDGRV 326
Db 1659 QGLRPGSEYTVSVVVALHDDMESQPLIGTQSTAIPTDPAKFTQVTPPTSLSAQWTPPNVQL 1718
QY 327 DGVR-----EKSDPHIKLOQAEERGVSIGVCANRY-----LANKEDGRLLASK----- 372
Db 1719 TGRVVRVTCKEKTGPMKEINLAPDSSSVVSGMLVATKYEVSVVALKD---TLTSRPAQ 1775
QY 373 -----CVTDEC-----FFERLESNNYNTYRS-----R 395
Db 1776 VVTTLENVSPRRARVTDATETITISWRKTETITIGFQVDVAVPANGQTPQRTIKPDVR 1835
QY 396 KYT-----SWYVA 403
Db 1836 SYITITGLQPGTDYKIYLYTLNDNARSPPVVIDASTAIDAPSNLRFLLATTPNSLLVSWQPP 1895
QY 404 LKRTGQY-----KLGSK-----TG--PGQKAILFL-----PM- 428
Db 1896 RARITGYIIKYEKPGSPRPVPRPGVTEATITGLEPGTEYTIYVIALKNNKSEPLI 1955
QY 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1956 GRKKTDELQVLTLPHPNLHGPEILDVPST 1985

RESULT 14

US-10-447-161-3
; Sequence 3, Application US/10447161
; Publication No. US20040023314A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Rong-fu
; TITLE OF INVENTION: Mutant Fibronectin and Tumor Metastasis
; FILE REFERENCE: HO-P02484US1
; CURRENT APPLICATION NUMBER: US/10/447,161
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/383,530
; PRIOR FILING DATE: 2002-05-28
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 2355
; TYPE: PRT
; ORGANISM: Human
US-10-447-161-3

Query Match 61.5%; Score 1474; DB 16; Length 2355;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;
QY 1 PTDLRFNIGPDMRTVMTWAPPSPSIDLTNFLVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60

Db 1270 PTDLFTNIGPDMRWVWAPPSPIDLTNLFVRYSPVKNEEDVAELSI SPSDNAVVLTNLL 1329

Qy 61 PCTEYVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120

Db 1330 PCTEYVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 1389

Qy 121 IRHPEHSGRPREDRVPHSRNSITLTNLTPTGTEYVSVSIVALNGREESPLLIQQOSTVSD 180

Db 1390 IRHPEHSGRPREDRVPHSRNSITLTNLTPTGTEYVSVSIVALNGREESPLLIQQOSTVSD 1449

Qy 181 VRDLVVAAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEVTPVGSKSTATISGLK 240

Db 1450 VRDLVVAAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEVTPVGSKSTATISGLK 1509

Qy 241 PGVDYTIIVYAVTGRGDSPASSKPISINRYTEIDKPS-----M 278

Db 1510 PGVDYTIIVYAVTGRGDSPASSKPISINRYTEIDKPSQMVQTDVQDINSIVKWLPSSEPV 1569

Qy 279 AAGSITTLTP----- 287

Db 1570 TGYRVTTTPKNGPGPKTKTAGPDQTEMTEIEGLQPTVEYVSVYAQNPSGESQPLVQTAV 1629

Qy 288 -----ALPEDGGSGAPP-PGHFKDPKRL 309

Db 1630 TNIDRPKGLAFTDVSISIKIAWESPOQVSRVRYTYSSPEDGIHELFPAPDGEEDTABEL 1689

Qy 310 YCKNGG-----FPLRIH-----PDGRV 326

Db 1690 QGLRGSEYTVSVVVALHDDMESQPLIGTQSTAI PAPTDLKFTQVTPTSLSAQWTPNVQL 1749

Qy 327 DGVR-----EKSDDPHIKLOQABERGVSIVKGCANRY-----LAWKEDGRLLASK- 372

Db 1750 TGYRVVTPKKTGPMKEINLAPDSSVVVSGLMVATKYEVSVAALKD---TLTSRPAQG 1806

Qy 373 -----CWTDEC-----PFRLESNNNTYRS-----R 395

Db 1807 VVTTLNVSPPRRARVTDATETITISWRKTETITGTFQVDVAVPANGOTPIQRTIKPDVR 1866

Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457

Db 1987 GRKKTDELQVLTLPHPNLHGPEILDVPST 2016

RESULT 15

US-09-961-403-1

Sequence 1, Application US/09961403

Publication No. US20030077589A1

GENERAL INFORMATION:

APPLICANT: HE-STUMPP, HOLGER

APPLICANT: HAENDLER, BERNARD

APPLICANT: KRAETZSCHMAR, JOERN

APPLICANT: KREFT, BERTHOLT

APPLICANT: WINTERHAGER, ELKE

APPLICANT: REGIDOR, PEDRO

APPLICANT: SCOTTI, SIMONE

TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS

FILE REFERENCE: SCH-1789

CURRENT APPLICATION NUMBER: US/09/961,403

CURRENT FILING DATE: 2001-09-25

NUMBER OF SEQ ID NOS: 15

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 2386

TYPE: PRN

ORGANISM: Homo sapiens

US-09-961-403-1

Query Match 61.5%; Score 1474; DB 10; Length 2386;

Best Local Similarity 46.8%; Pred. No. 1.7e-109;

Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLFTNIGPDMRWVWAPPSPIDLTNLFVRYSPVKNEEDVAELSI SPSDNAVVLTNLL 60

Db 1270 PTDLFTNIGPDMRWVWAPPSPIDLTNLFVRYSPVKNEEDVAELSI SPSDNAVVLTNLL 1329

Qy 61 PCTEYVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120

Db 1330 PCTEYVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 1389

Qy 121 IRHPEHSGRPREDRVPHSRNSITLTNLTPTGTEYVSVSIVALNGREESPLLIQQOSTVSD 180

Db 1390 IRHPEHSGRPREDRVPHSRNSITLTNLTPTGTEYVSVSIVALNGREESPLLIQQOSTVSD 1449

Qy 181 VRDLVVAAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEVTPVGSKSTATISGLK 240

Db 1450 VRDLVVAAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEVTPVGSKSTATISGLK 1509

Qy 241 PGVDYTIIVYAVTGRGDSPASSKPISINRYTEIDKPS-----M 278

Db 1510 PGVDYTIIVYAVTGRGDSPASSKPISINRYTEIDKPSQMVQTDVQDINSIVKWLPSSEPV 1569

Qy 279 AAGSITTLTP----- 287

Db 1570 TGYRVTTTPKNGPGPKTKTAGPDQTEMTEIEGLQPTVEYVSVYAQNPSGESQPLVQTAV 1629

Qy 288 -----ALPEDGGSGAPP-PGHFKDPKRL 309

Db 1630 TNIDRPKGLAFTDVSISIKIAWESPOQVSRVRYTYSSPEDGIHELFPAPDGEEDTABEL 1689

Qy 310 YCKNGG-----FPLRIH-----PDGRV 326

Db 1690 QGLRGSEYTVSVVVALHDDMESQPLIGTQSTAI PAPTDLKFTQVTPTSLSAQWTPNVQL 1749

Qy 327 DGVR-----EKSDDPHIKLOQABERGVSIVKGCANRY-----LAWKEDGRLLASK- 372

Db 1750 TGYRVVTPKKTGPMKEINLAPDSSVVVSGLMVATKYEVSVAALKD---TLTSRPAQG 1806

Qy 373 -----CWTDEC-----PFRLESNNNTYRS-----R 395

Db 1807 VVTTLNVSPPRRARVTDATETITISWRKTETITGTFQVDVAVPANGOTPIQRTIKPDVR 1866

Qy 396 KVT----- 403

Db 1867 SYTITGLQPGTDYKIYLYTLNDNARSSPVVIDASTAIDAPSNLRLFLATTPNSLLVSWQPP 1926

Qy 404 LKRTQY-----KLGSK-----TG--PGOKAILFL-----PM- 428

Db 1927 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIIVIALKNNQKSEPLI 1986

Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457

Db 1987 GRKKTDELQVLTLPHPNLHGPEILDVPST 2016

Search completed: May 3, 2004, 13:20:49

Job time : 59.7263 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:06:19 ; Search time 19.9005 Seconds
(without alignments)
2208.970 Million cell updates/sec

Title: US-09-775-964-5
Perfect score: 2398
Sequence: 1 PTDLRTNIGPDMRTWAP.....LVTLPHPNLHGPEILDVPST 457
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1474	61.5	2386	1 FNBU	fibronectin precu
2	1396	58.2	2265	1 FNBU	fibronectin - bovi
3	1345	56.1	2477	2 S14428	fibronectin precu
4	1018.5	42.5	2481	2 A43908	fibronectin - Afri
5	848	35.4	273	2 A28512	fibronectin - chic
6	824.5	34.4	210	2 A32398	basic fibroblast g
7	812	33.9	157	1 GKBOB	basic fibroblast g
8	792.5	32.0	154	2 A31674	basic fibroblast g
9	777.5	32.4	154	2 C37360	basic fibroblast g
10	768	32.0	189	2 S71465	fibronectin - chic
11	764	31.9	146	1 S00185	basic fibroblast g
12	754.5	31.5	164	2 S31622	basic fibroblast g
13	753	31.4	189	2 A48834	basic fibroblast g
14	738	30.8	137	2 I46711	fibroblast growth
15	681	28.4	155	1 A40117	basic fibroblast g
16	613.5	25.6	1020	2 A29355	fibronectin - chic
17	467	19.5	125	2 A32484	basic fibroblast g
18	425.5	17.7	155	1 A60721	acidic fibroblast
19	417.5	17.4	155	2 A60130	acidic fibroblast
20	416.5	17.4	155	1 A33665	acidic fibroblast
21	411.5	17.2	155	2 S04147	acidic fibroblast
22	411.5	17.2	155	2 D37360	acidic fibroblast
23	403.5	16.8	152	2 JH0476	acidic fibroblast
24	402.5	16.8	155	2 JW0055	acidic fibroblast
25	400.5	16.7	155	1 GKBOA	acidic fibroblast
26	378.5	15.8	1356	2 A45445	janusin precursor
27	358	14.9	1353	1 JH0675	restricin precurs
28	338.5	14.1	1746	1 S19694	tenascin precursor
29	334.5	13.9	2201	2 A32160	tenascin-C - human

30	333.5	13.9	929	2	I51027	type XII collagen
31	329	13.7	1810	1	A32230	tenascin precursor
32	327	13.6	2019	1	JQ1322	tenascin precursor
33	309	12.9	3124	2	A40020	collagen alpha 1(X
34	302.5	12.6	4135	2	T42629	tenascin-X - bovin
35	297.5	12.4	843	2	A40970	undulin 1 - human
36	290.5	12.1	1857	2	S31212	collagen alpha 1(X
37	290.5	12.1	1888	2	S78476	collagen alpha 1(X
38	289.5	12.1	1747	2	A45974	cytotactin - chick
39	280.5	11.7	933	2	A31930	tenascin-X precurs
40	275.5	11.5	3566	1	A40701	tenascin-X - mouse
41	274	11.4	860	2	I48839	probable tenascin
42	274	11.4	4006	2	T09070	collagen alpha 1(V
43	272	11.3	2944	2	A54849	fibroblast growth
44	270.5	11.3	206	1	TVHUHS	fibroblast growth
45	266	11.1	194	2	I50710	fibroblast growth

ALIGNMENTS

RESULT 1

FNBU

fibronectin precursor [validated] - human

N;Alternate names: fibronectin splice form ED-A

C;Species: Homo sapiens (man)

C;Date: 27-Nov-1985 #sequence revision 31-Mar-1993 #text change 08-Dec-2000

C;Accession: A26460; A26284; S03917; A24854; A24476; A91008; A93529; A21011; A90495; A2

R;Dean, D.C.; Bowlus, C.L.; Bourgeois, S.

Proc. Natl. Acad. Sci. U.S.A. 84, 1876-1880, 1987

A;Title: Cloning and analysis of the promoter region of the human fibronectin gene.

A;Reference number: A26460; MUID:87175578; PMID:3031656

A;Accession: A26460

A;Molecule type: DNA

A;Residues: 1-49 <DEA>

A;Cross-references: GB:M15801; NID:9182686; PIDN:AAA53376.1; PID:G553293

R;Oldberg, A.; Ruoslahti, E.

J. Biol. Chem. 261, 2113-2116, 1986

A;Title: Evolution of the fibronectin gene.

A;Reference number: A26284; MUID:86111901; PMID:3003095

A;Accession: A26284

A;Molecule type: DNA

A;Residues: 1447-1540 <OLD>

A;Cross-references: GB:M12549; NID:9182688

A;Note: the authors translated the codon TTC for residue 1494 as Glu

R;Paolella, G.; Henschcliffe, C.; Sebastio, G.; Baralle, F.E.

Nucleic Acids Res. 16, 3545-3557, 1988

A;Title: Sequence analysis and in vivo expression show that alternative splicing of ED-

A;Reference number: S00848; MUID:88233940; PMID:3375063

A;Accession: S03917

A;Molecule type: DNA

A;Residues: 1594-1767, 'V', 1769-1783 <PAO>

A;Cross-references: EMBL:X07118; NID:931402

A;Note: the authors translated the codon AAC for residue 1631 as Asp

R;Vibe-Pedersen, K.; Magnusson, S.; Baralle, F.E.

FEBS Lett. 207, 287-291, 1986

A;Title: Donor and acceptor splice signals within an exon of the human fibronectin gene

A;Reference number: A24854; MUID:87030929; PMID:3770201

A;Accession: A24854

A;Molecule type: DNA

A;Residues: 1992-2147 <VIB>

A;Cross-references: GB:X04530; NID:931436

R;Gutman, A.; Yamada, K.M.; Kornblitt, A.

FEBS Lett. 207, 145-148, 1986

A;Title: Human fibronectin is synthesized as a pre-propolypeptide.

A;Reference number: A24476; MUID:87030890; PMID:3770189

A;Accession: A24476

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-14, 'O', 16-38 <GUT>

R;Kornblitt, A.R.; Umezawa, K.; Vibe-Pedersen, K.; Baralle, F.E.

EMBO J. 4, 1755-1759, 1985

A;Title: Primary structure of human fibronectin: differential splicing may generate at

A;Reference number: A91008; MUID:85284965; PMID:2992939
A;Accession: A91008
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 32-1344;1346-2080;2112-2386 <KOR>
A;Cross-references: GB:X02761
R;Kornblitt, A.R.; Vibe-Pedersen, K.; Baralle, F.E.
Nucleic Acids Res. 12, 5853-5868, 1984
A;Title: Human fibronectin: cell specific alternative mRNA splicing generates polypeptide
A;Reference number: A93529; MUID:84272258; PMID:6462919
A;Accession: A93529
A;Molecule type: mRNA
A;Residues: 973-2080;2112-2386 <KOR2>
A;Cross-references: GB:X00739
R;Oldberg, A.; Linney, E.; Ruoslahti, E.
J. Biol. Chem. 258, 10193-10196, 1983
A;Title: Molecular cloning and nucleotide sequence of a cDNA clone coding for the cell a
A;Reference number: A21011; MUID:83290929; PMID:6688418
A;Accession: A21011
A;Molecule type: mRNA
A;Residues: 1434-1537 <OL2>
A;Cross-references: GB:K00055; NID:gl82680; PIDN:AAA52459.1; PID:gl82683
R;Bernard, M.F.; Kolbe, M.; Weil, D.; Chu, M.L.
Biochemistry 24, 2698-2704, 1985
A;Title: Human cellular fibronectin: comparison of the carboxyl-terminal portion with ra
A;Reference number: A90495; MUID:85280409; PMID:2992573
A;Accession: A90495
A;Molecule type: mRNA
A;Residues: 1594-2386 <BER>
A;Cross-references: GB:M10905; NID:gl82696; PIDN:AAA52462.1; PID:gl82697
R;Umezawa, K.; Kornblitt, A.R.; Baralle, F.E.
FEBS Lett. 186, 31-34, 1985
A;Title: Isolation and characterization of cDNA clones for human liver fibronectin.
A;Reference number: A22245; MUID:85231203; PMID:2989004
A;Accession: A22245
A;Molecule type: mRNA
A;Residues: 1948-2067 <UME>
A;Cross-references: GB:M27589; NID:gl82705; PIDN:AAA52465.1; PID:gl82706
A;Accession: B22245
A;Molecule type: mRNA
A;Residues: 1975-1991;2017-2039 <UM2>
A;Cross-references: GB:M27590
R;Sekiguchi, K.; Klos, A.M.; Kurachi, K.; Yoshitake, S.; Hakomori, S.
Biochemistry 25, 4936-4941, 1986
A;Title: Human liver fibronectin complementary DNAs: identification of two different mes
A;Reference number: 152394; MUID:87026578; PMID:3021206
A;Accession: I65273
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1978-1990;2017-2039 <SEK>
A;Cross-references: GB:M14060; NID:gl82701; PIDN:AAA52464.1; PID:gl82704
R;Kornblitt, A.R.; Vibe-Pedersen, K.; Baralle, F.E.
Proc. Natl. Acad. Sci. U.S.A. 80, 3218-3222, 1983
A;Title: Isolation and characterization of cDNA clones for human and bovine fibronectins
A;Reference number: A21165; MUID:83221567; PMID:6304699
A;Accession: A21165
A;Molecule type: mRNA
A;Residues: 2291-2386 <KOR3>
A;Cross-references: GB:K00799; NID:gl82681; PIDN:AAA52460.1; PID:gl82684
R;Garcia-Pardo, A.; Pearlstein, E.; Frangione, B.
J. Biol. Chem. 258, 12670-12674, 1983
A;Title: Primary structure of human plasma fibronectin.
A;Reference number: A92398; MUID:84032463; PMID:6630202
A;Accession: A92398
A;Molecule type: protein
A;Residues: 32-47,'C',49-51,'S',53-72,'A',74-290 <GAR1>
R;Garcia-Pardo, A.; Gold, L.I.
Arch. Biochem. Biophys. 304, 181-188, 1993
A;Title: Further characterization of the binding of fibronectin to gelatin reveals the F
A;Reference number: S34791; MUID:93312001; PMID:8323285
A;Accession: S34791
A;Molecule type: protein
A;Residues: 291-300;551-560 <GAR2>

R;Griffin, C.A.; Calaycay, J.; Shively, J.E.; Smith, R.L.
Thromb. Res. 43, 469-477, 1986
A;Title: Two plasma fibronectin fragments with different gelatin-binding properties.
A;Reference number: A60904; MUID:87019725; PMID:3532418
A;Accession: A60904
A;Molecule type: protein
A;Residues: 293-301 <GRI>
R;Calaycay, J.; Pande, H.; Lee, T.; Borsi, L.; Siri, A.; Shively, J.E.; Zardi, L.
J. Biol. Chem. 260, 12136-12141, 1985
A;Title: Primary structure of a DNA- and heparin-binding domain (domain III) in human pl
A;Reference number: A23901; MUID:86008277; PMID:3900070
A;Accession: A23901
A;Molecule type: protein
A;Residues: 616-677,'Q',679-703,'PT' <CAL>
R;Pierschbacher, M.D.; Ruoslahti, E.; Sundelin, J.; Lind, P.; Peterson, P.A.
J. Biol. Chem. 257, 9593-9597, 1982
A;Title: The cell attachment domain of fibronectin. Determination of the primary structu
A;Reference number: A92386; MUID:82265604; PMID:7050098
A;Accession: A92386
A;Molecule type: protein
A;Residues: 1441-1548 <PIE>
A;Note: residues 1524-1527 are responsible for the cell-binding activity
R;Garcia-Pardo, A.; Rostagno, A.; Frangione, B.
Biochem. J. 241, 923-928, 1987
A;Title: Primary structure of human plasma fibronectin. Characterization of a 38 kDa dom
A;Reference number: A32517; MUID:87241275; PMID:3593230
A;Accession: A32517
A;Molecule type: protein
A;Residues: 1589-1630,'T',1722-2058 <GAR3>
R;Tresselt, T.; McCarthy, J.B.; Calaycay, J.; Lee, T.D.; Legesse, K.; Shively, J.E.; Pand
Biochem. J. 274, 731-738, 1991
A;Title: Human plasma fibronectin. Demonstration of structural differences between the A
A;Reference number: S14357; MUID:91190085; PMID:2012601
A;Accession: S14357
A;Molecule type: protein
A;Residues: 1614-1630,'T',1722-2081,2113-2244 <TRE>
R;Garcia-Pardo, A.; Pearlstein, E.; Frangione, B.
J. Biol. Chem. 260, 10320-10325, 1985
A;Title: Primary structure of human plasma fibronectin. Characterization of a 31,000-dal
A;Reference number: A23891; MUID:85261459; PMID:4019516
A;Accession: A23891
A;Molecule type: protein
A;Residues: 2071-2080;2112-2356 <GAR4>
C;Comment: The extra domain and connecting strand 3 are subject to developmental and tis
C;Comment: The cellular and plasma fibronectins are high molecular weight glycoproteins,
ation, and transformation.
C;Genetics:
A;Gene: GDB:FNI
A;Cross-references: GDB:119135; OMIM:135600
A;Map position: 2q34-q34
A;Introns: 49/3; 1266/1; 1357/1; 1447/1; 1487/1; 1541/1; 1631/1; 1721/1; 1991/1; 2145/1
C;Superfamily: fibronectin; fibronectin type I repeat homology <F1>
C;Keywords: acute phase; alternative splicing; cell adhesion; collagen binding; duplicat
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-31/Domain: propeptide #status predicted <PRO>
F;32-2386/Product: fibronectin #status experimental <MAT>
F;52-272/Domain: fibrin and heparin binding <PHB>
F;52-272/Domain: fibronectin type I repeat homology <F1>
F;97-135/Domain: fibronectin type I repeat homology <F2>
F;141-179/Domain: fibronectin type I repeat homology <F3>
F;186-225/Domain: fibronectin type I repeat homology <F4>
F;231-270/Domain: fibronectin type I repeat homology <F5>
F;308-608/Domain: collagen binding <CBR>
F;308-342/Domain: fibronectin type I repeat homology <F6>
F;360-401/Domain: fibronectin type II repeat homology <F1>
F;420-461/Domain: fibronectin type II repeat homology <F2>
F;470-508/Domain: fibronectin type I repeat homology <F7>
F;518-555/Domain: fibronectin type I repeat homology <F8>
F;561-599/Domain: fibronectin type I repeat homology <F9>
F;609-692/Domain: fibronectin type III repeat homology <3FA>
F;616-706/Domain: heparin binding <HPB>
F;719-801/Domain: fibronectin type III repeat homology <3FB>
F;810-891/Domain: fibronectin type III repeat homology <3FC>

F;906-988/Domain: fibronectin type III repeat homology <3PD>	
F;996-1077/Domain: fibronectin type III repeat homology <3FE>	
F;1086-1164/Domain: fibronectin type III repeat homology <3FF>	
F;1173-1258/Domain: fibronectin type III repeat homology <3FG>	
F;1266-1349/Domain: fibronectin type III repeat homology <3FH>	
Query Match	61.5%; Score 1474; DB 1; Length 2386;
Best Local Similarity	46.8%; Pred. No. 3.7e-97;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;	
Qy	1 PTDLFTNIGPDTRVTVWAPPSPIDLTNLFVRVSPVKNEDVAELSTSPSDNAVLTNLL 60
Db	1270 PTDLFTNIGPDTRVTVWAPPSPIDLTNLFVRVSPVKNEDVAELSTSPSDNAVLTNLL 1329
Qy	61 PGTEYVVSVSVEQHESTELRGKQTKGLDPTGIDFSDITANSFTVHWTAPRATITGYR 120
Db	1330 PGTEYVVSVSVEQHESTELRGKQTKGLDPTGIDFSDITANSFTVHWTAPRATITGYR 1389
Qy	121 IRHPEHPSGRPREDRVPHSRNSITLTNLTGTEYVVSIVVALNGREESPLLIGQSTVSD 180
Db	1390 IRHPEHPSGRPREDRVPHSRNSITLTNLTGTEYVVSIVVALNGREESPLLIGQSTVSD 1449
Qy	181 VPRDLEVVAAATPSTLLISWAPAVTVVYRITYGETGNSPVQEFVPGSKSTATISGLK 240
Db	1450 VPRDLEVVAAATPSTLLISWAPAVTVVYRITYGETGNSPVQEFVPGSKSTATISGLK 1509
Qy	241 PGVDYTTIVAAVTGRGDSPASSKPIISINRYTEIDKPS-----M 278
Db	1510 PGVDYTTIVAAVTGRGDSPASSKPIISINRYTEIDKPSQMQVTDVQDNSISVKWLPSSSPV 1569
Qy	279 AAGSITLPL----- 287
Db	1570 TGRVTVTTTPKNGPGTKTAGPDQTEMTIEGLQPTVEYVVSVAQNPGSESQPLVQTAV 1629
Qy	288 -----ALPEDGSGAFP-PGHFKQPKRL 309
Db	1630 TNIDRPKGLAFTVDVDSIKIAMESPGQVSRVRYTVSSPEDGHELFPAAPDGEEDTAEL 1689
Qy	310 YCKNGG-----FFLRIH-----PDGRV 326
Db	1690 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIPTAPDLKFTQVTPTSLSAQWTPPNVL 1749
Qy	327 DQVR-----EKSDPHIKLOQAEERGVSIVKVCANRY-----LAMKEDGRLLASK---- 372
Db	1750 TGYRVRVTPKEKTGPMKEINLAPDSSSVVWGLMVAATKYEVSVALKD---TLTSRPAQG 1806
Qy	373 -----CVTDEC-----FFERLESNNYTYRS-----R 395
Db	1807 VVTTLENVSPRRARVTDATETTTISWRKTETITGQVDVAPANGQTPQIRTIKPDVR 1866
Qy	396 KYT-----GWYVA 403
Db	1867 SYTITGLQGTQDKYIYLTINDNARSPVVIDASTAIDAPSNRFLATTNSLLVSWQPP 1926
Qy	404 LKRTGQY-----KLGSK-----TG-PQOKAILFL-----PM- 428
Db	1927 RARITGVIIKVEKPGSPREWPVRPRPGVTEATITGLEPGTEYTVIVALKNKKQSEPLI 1986
Qy	429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db	1987 GRXKTDLPQLVTLPHPNLHGPEILDVPST 2016

RESULT 2

FNBO

fibronectin - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 31-Dec-1988 #sequence revision 31-Dec-1988 #text_change 20-Oct-2000

C;Accession: A26452; B21165; A23292

R;Skorstengaard, K.; Jensen, M.S.; Sahl, P.; Petersen, T.E.; Magnusson, S.

Eur. J. Biochem. 161, 441-453, 1986

A;Title: Complete primary structure of bovine plasma fibronectin.

A;Reference number: A26452; MUID:87054047; PMID:3780752

A;Accession: A26452
A;Molecule type: protein
A;Residues: 1-2265 <SKO>
R;Kornblitt, A.R.; Vibe-Pedersen, K.; Baralle, F.E.
Proc. Natl. Acad. Sci. U.S.A. 80, 3218-3222, 1983
A;Title: Isolation and characterization of cDNA clones for human and bovine fibronectin
A;Reference number: A21165; MUID:83221567; PMID:6304699
A;Accession: B21165
A;Molecule type: mRNA
A;Residues: 2170-2265 <KOR>
A;Cross-references: GB:K00800; NID:G163055; PIDN:AAA30521.2; PID:G5713323
R;Petersen, T.E.; Thogersen, H.C.; Skorstengaard, K.; Vibe-Pedersen, K.; Sahl, P.; Sott
Proc. Natl. Acad. Sci. U.S.A. 80, 137-141, 1983
A;Title: Partial primary structure of bovine plasma fibronectin: three types of interna
A;Reference number: A23292; MUID:83117805; PMID:6218503
A;Accession: A23292
A;Molecule type: protein
A;Residues: 1-16, C, 18-20, 'S', 22-432;447-463;1367-1517;1567-1673;2062-2176, 'N', 2178-22
C;Comment: Cys-1201 and Cys-2015 have free sulphydryl groups.
C;Comment: The plasma fibronectin molecule consists of two chains, which are connected i
C;Comment: Fibronectins bind cell surfaces and various compounds including collagen, fi
aling, and maintenance of cell shape.
C;Comment: Plasma fibronectin is synthesized by hepatocytes.
C;Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II rep
C;Keywords: acute phase; alternative splicing; collagen binding; duplication; extracell
F;21-241/Domain: fibrin and heparin binding <FBR>
F;66-104/Domain: fibronectin type I repeat homology <1F1>
F;110-148/Domain: fibronectin type I repeat homology <1F2>
F;135-194/Domain: fibronectin type I repeat homology <1F3>
F;200-239/Domain: fibronectin type I repeat homology <1F4>
F;277-577/Domain: collagen binding <CBR>
F;329-370/Domain: fibronectin type I repeat homology <1F6>
F;389-430/Domain: fibronectin type II repeat homology <2F1>
F;439-477/Domain: fibronectin type II repeat homology <2F2>
F;487-524/Domain: fibronectin type I repeat homology <1F7>
F;530-568/Domain: fibronectin type I repeat homology <1F8>
F;578-661/Domain: fibronectin type III repeat homology <1F9>
F;688-770/Domain: fibronectin type III repeat homology <FN3A>
F;779-860/Domain: fibronectin type III repeat homology <FN3B>
F;875-957/Domain: fibronectin type III repeat homology <FN3C>
F;965-1046/Domain: fibronectin type III repeat homology <FN3E>
F;1055-1134/Domain: fibronectin type III repeat homology <FN3F>
F;1142-1227/Domain: fibronectin type III repeat homology <FN3G>
F;1235-1318/Domain: fibronectin type III repeat homology <FN3H>
F;1326-1404/Domain: fibronectin type III repeat homology <GN3I>
F;1410-1517/Domain: cell attachment <CAD>
F;1416-1502/Domain: fibronectin type III repeat homology <FN3J>
F;1493-1495/Region: cell attachment (R-G-D) motif
F;1510-1592/Domain: fibronectin type III repeat homology <FN3K>
F;1600-1870/Domain: heparin binding <HB2>
F;1600-1682/Domain: fibronectin type III repeat homology <FN3L>
F;1692-1773/Domain: fibronectin type III repeat homology <FN3M>
F;1781-1863/Domain: fibronectin type III repeat homology <FN3N>
F;1970-1972/Region: cell attachment (R-G-D) motif
F;1982-2062/Domain: fibronectin type III repeat homology <FN3O>
F;1985-2216/Domain: fibrin binding <FB2>
F;2085-2124/Domain: fibronectin type I repeat homology <1F10>
F;2130-2167/Domain: fibronectin type I repeat homology <1F11>
F;2174-2209/Domain: fibronectin type I repeat homology <1F12>
F;1/Modified site: pyroglutamate (Gln) (interchain to Lys N6-amino of fibrin) #status experimental
F;3/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino of fibrin) #status experim
F;399,497,511,846,976,1213,1987/Binding site: carbohydrate (Asn) (covalent) #status absent
F;1205,1692/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;1943,1944/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;2246/Diulfide bonds: interchain (to 2250) #status predicted
F;2250/Diulfide bonds: interchain (to 2246) #status predicted
F;2263/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match

58.2%; Score 1396; DB 1; Length 2265;

Best Local Similarity 48.6%; Pred. No. 1.4e-91;		Matches 327; Conservative 38; Mismatches 76; Indels 232; Gaps 17;	
QY	1	PTDLRTNIGDTRMTVWAPPSSIDLTNLFVYSPVKNEEDVAELISPSDNAVLTNLL	60
Db	1239	PTDLRTNIGDTRMTVWAPPSSIELTNLIVRSPVKNEEDVAELISPSDNAVLTNLL	1298
QY	61	PGTEYVSVSSVVEQHESTPLGRQKTGLDSTGIDFSDITANSTFTVHWIAPRAITGYR	120
Db	1299	PGTEYVSVSVVEQHESTPLGRQKTALDPSGIDFSDITANSTFTVHWIAPRAITGYR	1358
QY	121	IRHPHFHSGRPREDRVPHSRNSITLTNLTGTEYVSVSIVALNGRESPLLIGQOSTVSD	180
Db	1359	IRHPENMGRRPREDRVPHSRNSITLTNLTGTEYVSVSIVALNKEESPLVQGOQSTVSD	1418
QY	181	VPRLVVAATPTSLLSISWDAPAVTVRYRITYGETGNSPQVQFTVPQSKSTATISGLK	240
Db	1419	VPRLVVAATPTSLLSISWDAPAVTVRYRITYGETGSSPVQVQFTVPQSKSTATISGLK	1478
QY	241	PGVDYITTVAVTGRGDSPASSKPIISINVTETIDKPSMA-----AGSITTLPA----	288
Db	1479	PGVDYITTVAVTGRGDSPASSKPIISINVTETIDKPSQMVQVTDVQDNSISVRWLFSPPV	1538
QY	289	-----LPEDGGSGAFPPGCHFDPK-----	313
Db	1539	TGVRVTTAPKNG-----PGPSKTKTVGPDQTEMTIEGLQPTVEYVSVYAQNQGESQP	1592
QY	314	-----	337
Db	1593	LVQTAVTTPAPTNLKFTQVTPTSLTQAOWTAPNVQLTGYRVRTP-----	1645
QY	338	KLQLOAEERGVSIGKVCANRY-----LAMEDGRLASK-----	375
Db	1646	EINLAPDSSVVVSGLMVATKVEVSVALKD---TLTSRPAQGVVTTLENVSPPRARVT	1702
QY	376	DEC-----	396
Db	1703	DAFETTTISWRTKTTETITGTFQVDAIPANGQTPQIORTIPDRVSYYITGLQPGTDYKIHL	1762
QY	397	YT-----	415
Db	1763	YTLNDNARSPVVIDASTAIDAPSINLRFIATTPNSLLVSWQPPRARITGYIITYKXPGPS	1822
QY	416	-----TG--PGQK---ATLFLPMSAAS-----	444
Db	1823	PREVVRPRPGVTEAITGLPGTETVTIQVIALKNNQKSEPLIGRKKKTDELQVLTPLPH	1882
QY	445	NLHGPEILDVPST 457	
Db	1883	NLHGPEILDVPST 1895	
RESULT 3			
S14428			
fibronectin precursor - rat			
C;Species: Rattus norvegicus (Norway rat)			
C;Date: 28-Oct-1994 #sequence revision 28-Oct-1994 #text change 20-Aug-1999			
C;Accession: S14428; S12455; A22319; S46203; S00459; A27252; I59049			
R;Hynes, R.O.			
submitted to the EMBL Data Library, July 1989			
A;Reference number: S14428			
A;Accession: S14428			
A;Molecule type: mRNA			
A;Residues: 1-2477 <N>			
A;Cross-references: EMBL:X15906; NID:g56163; PIDN:CAA34020.1; PID:g56164			
R;Schwarzbauer, J.E.; Patel, R.S.; Fonda, D.; Hynes, R.O.			
EMBO J. 6, 2573-2580, 1987			
A;Title: Multiple sites of alternative splicing of the rat fibronectin gene transcript.			
A;Reference number: S12455; MUID:88054951; PMID:2445560			
A;Accession: S12455			
A;Status: nucleic acid sequence not shown			
A;Molecule type: mRNA			
A;Residues: 609-1810,'T',1812-2283 <SCH>			

A;Cross-references: EMBL:X15906		R;Tamkun, J.W.; Schwarzbauer, J.E.; Hynes, R.O.	
Proc. Natl. Acad. Sci. U.S.A. 81, 5140-5144, 1984		A;Title: A single rat fibronectin gene generates three different mRNAs by alternative sp	
A;Reference number: A22319; MUID:84298097; PMID:6089177		A;Accession: A22319	
A;Molecule type: DNA		A;Residues: 2052-2237 <TAM>	
R;Falkenberg, C.; Enghild, J.J.; Thøgersen, I.B.; Salvesen, G.; Akerstrøm, B.		Biochem. J. 301, 745-751, 1994	
A;Title: Isolation and characterization of fibronectin-alpha (1)-microglobulin complex in		A;Reference number: S46203; MUID:94330948; PMID:7519849	
A;Accession: S46203		A;Status: preliminary	
A;Molecule type: protein		A;Residues: 1183-1192,'G'N',1268,'P',1270-1271,'D',1273,'CF',1276,'PY',1385-1399 <FAL>	
R;Facell, R.S.; Odermatt, E.; Schwarzbauer, J.E.; Hynes, R.O.		EMBO J. 6, 2565-2572, 1987	
A;Title: Organization of the fibronectin gene provides evidence for exon shuffling durin		A;Reference number: S00459; MUID:88054950; PMID:3119323	
A;Accession: S00459		A;Molecule type: DNA	
A;Residues: 1-139;2382-2477 <PAT>		A;Cross-references: EMBL:X05831	
A;Note: the authors translated the codon CCT for residues 51 and 94 as Ala		R;Schwarzbauer, J.E.; Tamkun, J.W.; Lemischka, I.R.; Hynes, R.O.	
Cell 35, 421-431, 1983		A;Title: Three different fibronectin mRNAs arise by alternative splicing within the codi	
A;Reference number: A27252; MUID:84082067; PMID:6317187		A;Accession: A27252	
A;Molecule type: mRNA		A;Residues: 1586-1720,'T',1722,1813-2477 <SC2>	
R;Odermatt, E.; Tamkun, J.W.; Hynes, R.O.		Proc. Natl. Acad. Sci. U.S.A. 82, 6571-6575, 1985	
A;Title: Repeating modular structure of the fibronectin gene: Relationship to protein st		A;Reference number: I59049; MUID:86016741; PMID:3863113	
A;Accession: I59049		A;Status: translated from GB/EMBL/DBDJ	
A;Molecule type: DNA		A;Residues: 1722-1810 <RS>	
A;Cross-references: GB:M11750; NID:g204164; PIDN:AAA41170.1; PID:g5554437		C;Genetics:	
A;Introns: 51/1; 94/1; 2416/3; 2454/3		C;Superfamily: fibronectin; fibronectin type I repeat homology <1F1>	
C;Keywords: alternative splicing; cell adhesion; collagen binding; disulfide bond; dupli		F;1-3/Domain: signal sequence #status predicted <SIG>	
F;33-2477/Product: fibronectin #status predicted <MAT>		F;53-88/Domain: fibronectin type I repeat homology <1F1>	
F;98-136/Domain: fibronectin type I repeat homology <1F2>		F;142-180/Domain: fibronectin type I repeat homology <1F3>	
F;187-226/Domain: fibronectin type I repeat homology <1F4>		F;232-271/Domain: fibronectin type I repeat homology <1F5>	
F;308-342/Domain: fibronectin type II repeat homology <2F1>		F;360-401/Domain: fibronectin type II repeat homology <2F2>	
F;420-461/Domain: fibronectin type II repeat homology <1F7>		F;470-508/Domain: fibronectin type I repeat homology <1F8>	
F;518-555/Domain: fibronectin type I repeat homology <1F9>		F;561-599/Domain: fibronectin type III repeat homology <FN3A>	
F;609-692/Domain: fibronectin type III repeat homology <FN3B>		F;718-800/Domain: fibronectin type III repeat homology <FN3C>	
F;809-890/Domain: fibronectin type III repeat homology <FN3D>		F;905-987/Domain: fibronectin type III repeat homology <FN3E>	
F;995-1076/Domain: fibronectin type III repeat homology <FN3F>		F;1085-1164/Domain: fibronectin type III repeat homology <FN3G>	
F;1172-1257/Domain: fibronectin type III repeat homology <FN3H>		F;1269-1348/Domain: fibronectin type III repeat homology <FN3I>	
F;1356-1439/Domain: fibronectin type III repeat homology <FN3J>		F;1447-1529/Domain: fibronectin type III repeat homology <FN3K>	
F;1537-1619/Domain: fibronectin type III repeat homology <FN3L>		F;1614-1616/Region: cell attachment (R-G-D) motif	
F;1631-1713/Domain: fibronectin type III repeat homology <FN3M>		F;1721-1803/Domain: fibronectin type III repeat homology <FN3N>	
F;1811-1893/Domain: fibronectin type III repeat homology <FN3O>			

F:1903-1984/Domain: fibronectin type III repeat homology <FN3O>
F:1992-2074/Domain: fibronectin type III repeat homology <FN3P>
F:2181-2183/Region: cell attachment (R-G-D) motif
F:2193-2273/Domain: fibronectin type III repeat homology <FN3Q>
F:2296-2335/Domain: fibronectin type I repeat homology <FN3O>
F:2341-2378/Domain: fibronectin type I repeat homology <FN3Q>
F:2385-2420/Domain: fibronectin type I repeat homology <FN3O>
F:53-79,77-88,98-126,124-136,142-170,168-180,187-216,214-226,232-261,259-271,308-335,333
368,2368-2378,2385-2411,2409-2420/Disulfide bonds: #status predicted
F:2458/Disulfide bonds: interchain (to 2462) #status predicted
F:2462/Disulfide bonds: interchain (to 2458) #status predicted

Query Match 56.1%; Score 1345; DB 2; Length 2477;
Best Local Similarity 42.4%; Pred. No. 7.1e-88;
Matches 318; Conservative 49; Mismatches 87; Indels 296; Gaps 18;

Qy 1 PTDLRFNIGDPTMRVTWAPPSPSIDLTNLFVRSYKNEEDVAELSLSPSDNAVLTNLL 60
Db 1360 PTDLRFNIGDPTMRVTWAPPSPSIDLTNLFVRSYKNEEDVAELSLSPSDNAVLTNLL 1419

Qy 61 PGTEYVSVSVSYEQHSTPLGRQKTKGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120
Db 1420 PGTEYVSVSVSYEQHSTPLGRQKTKGLDSTGIDFSDITANSFTVHVIAPRATITGYR 1479

Qy 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTYVSVSVVAVNGRESPLLIQOQSTVSD 180
Db 1480 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTYVSVSVVAVNGRESPLLIQOQSTVSD 1539

Qy 181 VPRDLVVAATPTSLISWDAPAVTVYRITTYGETGNSPVQEFVPGSKSTATISGLK 240
Db 1840 VPRDLVVAATPTSLISWDAPAVTVYRITTYGETGNSPVQEFVPGSKSTATISGLK 240

Qy 241 PGVDYTTVAVTGRGDSPASSKPKISINRYTEIDKPS-----M 278
Db 1600 PGVDYTTVAVTGRGDSPASSKPKISINRYTEIDKPSQMVTDVQDMSISVRWLPSTSPV 1559

Qy 279 AAGSITLPL----- 287
Db 1660 TGYRVTTAPKNGLGFTKTSQTSVSPDQTEMTIEGLQPTVEYVSVVAVNQRNGESPLVQTAV 1719

Qy 288 -----ALPEDGGSGAFP-PGHFKDPKRL 309
Db 1720 TNIDRPKGLAFTDVDVDSIKIAMESPOGVSRYRTVYSSPEDGHELFAPDGDDETAEL 1779

Qy 310 Y-----CKNGGF-----FLRIH-----PDGRV 326
Db 1780 HGLRPSGSEYTVSVVALHGEMESQPLIGVQSTAIAPATNLKFTQVSPITLTAQWTAPSKVL 1839

Qy 327 DGVR-----EKSDPHIKLOQAERGWSIKGVCANRY-----LAMKEDGRLLASK--- 372
Db 1840 TGYRVVTPKEKTGPMKEINLSPDSTSVIVSGMLVATKYEVSVVALKD---TLTSRPAQG 1896

Qy 373 -----CVTDEC-----FFERLESNNY----- 389
Db 1897 VVTTLENVSPRRARVTDATETITISWRTKTEITITGFDVAIPANGQTPVQRTISPDVR 1956

Qy 390 -----NTVRSRYT-----SHYVA 403
Db 1957 SYTITGLQPDGYKHLYTLNDNARSPPVIDASTAIDAPSNLRFLLTTNSLLVSWAP 2016

Qy 404 LKRTGOY-----KLGSK-----TG-PGQKAILFL-----PM- 428
Db 2017 RARITGVIIKYKPGSPRPVPRPGVTEATITGLEPGTEYTVIYVIALKNNKQKSEPLI 2076

Qy 429 -SAASDELQVLTLPNHPNHPGPEILLVDPST 457
Db 2077 GRKKTDELQVLTLPNHPNHPGPEILLVDPST 2106

RESULT 4
A43908
fibronectin - African clawed frog
C:Species: Xenopus laevis (African clawed frog)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 12-Feb-1999
C:Accession: A43908
R:DeSimone, D.W.; Norton, P.A.; Hynes, R.O.
Dev. Biol. 149, 357-369, 1992
A:Title: Identification and characterization of alternatively spliced fibronectin mRNAs
A:Reference number: A43908; MUID:92111942; PMID:1730390
A:Accession: A43908
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2481 <DES>
A:Cross-references: GB:M77820
A:Note: sequences extracted from NCBI backbone (NCBIP:77473)
C:Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II rep
C:Keywords: duplication; extracellular matrix; glycoprotein; heterodimer
F:55-90/Domain: fibronectin type I repeat homology <FN1>
F:100-138/Domain: fibronectin type I repeat homology <FN2>
F:144-182/Domain: fibronectin type I repeat homology <FN3>
F:189-228/Domain: fibronectin type I repeat homology <FN4>
F:234-273/Domain: fibronectin type I repeat homology <FN5>
F:309-343/Domain: fibronectin type I repeat homology <FN6>
F:361-402/Domain: fibronectin type II repeat homology <FN2F1>
F:421-462/Domain: fibronectin type II repeat homology <FN2F2>
F:471-509/Domain: fibronectin type I repeat homology <FN7>
F:519-556/Domain: fibronectin type I repeat homology <FN8>
F:562-600/Domain: fibronectin type I repeat homology <FN9>
F:610-693/Domain: fibronectin type III repeat homology <FN3A>
F:719-801/Domain: fibronectin type III repeat homology <FN3B>
F:810-891/Domain: fibronectin type III repeat homology <FN3C>
F:906-988/Domain: fibronectin type III repeat homology <FN3D>
F:996-1077/Domain: fibronectin type III repeat homology <FN3E>
F:1086-1165/Domain: fibronectin type III repeat homology <FN3F>
F:1173-1258/Domain: fibronectin type III repeat homology <FN3G>
F:1266-1349/Domain: fibronectin type III repeat homology <FN3H>
F:1357-1440/Domain: fibronectin type III repeat homology <FN3I>
F:1448-1530/Domain: fibronectin type III repeat homology <FN3J>
F:1538-1620/Domain: fibronectin type III repeat homology <FN3K>
F:1615-1617/Region: cell attachment (R-G-D) motif
F:1632-1714/Domain: fibronectin type III repeat homology <FN3L>
F:1722-1804/Domain: fibronectin type III repeat homology <FN3M>
F:1812-1894/Domain: fibronectin type III repeat homology <FN3N>
F:1904-1985/Domain: fibronectin type III repeat homology <FN3O>
F:1993-2075/Domain: fibronectin type III repeat homology <FN3P>
F:2197-2277/Domain: fibronectin type III repeat homology <FN3Q>
F:2301-2340/Domain: fibronectin type I repeat homology <FN10>
F:2346-2383/Domain: fibronectin type I repeat homology <FN11>
F:2390-2425/Domain: fibronectin type I repeat homology <FN12>
F:55-81,79-90,100-128,126-138,144-172,170-182,189-218,216-228,234-263,261-273,309-336,3
2373,2371-2383,2390-2416,2414-2425/Disulfide bonds: #status predicted
F:2459/Disulfide bonds: interchain (to 2463) #status predicted
F:2463/Disulfide bonds: interchain (to 2459) #status predicted

Query Match 42.5%; Score 1018.5; DB 2; Length 2481;
Best Local Similarity 68.7%; Pred. No. 1.9e-64;
Matches 195; Conservative 43; Mismatches 45; Indels 1; Gaps 1;

Qy 1 PTDLRFNIGDPTMRVTWAPPSPSIDLTNLFVRSYKNEEDVAELSLSPSDNAVLTNLL 60
Db 1361 PTDLRFNIGDPTMRVTWAPPSPSIDLTNLFVRSYKNEEDVAELSLSPSDNAVLTNLL 1420

Qy 61 PGTEYVSVSVSYEQHSTPLGRQKTKGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120
Db 1421 PGTEYVSVSVSYEQHSTPLGRQKTKGLDSTGIDFSDITANSFTVHVIAPRATITGYR 1480

Qy 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTYVSVSVVAVNGRESPLLIQOQSTVSD 180
Db 1481 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTYVSVSVVAVNGRESPLLIQOQSTVSD 1540

Qy 181 VPRDLVVAATPTSLISWDAPAVTVYRITTYGETGNSPVQEFVPGSKSTATISGLK 240
Db 1541 VPRDLVVAATPTSLISWDAPAVTVYRITTYGETGNSPVQEFVPGSKSTATISGLK 240

Qy 241 PGVDYTTVAVTGRGDSPASSKPKISINRYTEIDKP-SMAAGSI 283
Db 241 PGVDYTTVAVTGRGDSPASSKPKISINRYTEIDKP-SMAAGSI 283

Db 1601 PGVSYTITVYAVTGRGDSPASSKPLTIIHKTDVDPIDMAVTDI 1644

RESULT 5

A28512

fibronectin - chicken (fragment)

C:Species: Gallus gallus (chicken)

C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 20-Aug-1999

C:Accession: A28512

R: Kubomura, S.; Obara, M.; Karasaki, Y.; Taniguchi, H.; Gotoh, S.; Tsuda, T.; Higashi, K.

Biochim. Biophys. Acta 910, 171-181, 1987

A>Title: Genetic analysis of the cell binding domain region of the chicken fibronectin g

A:Reference number: A28512; MUID:88050950; PMID:2823899

A:Accession: A28512

A:Molecule type: DNA

A:Residues: 1-273 <KUB>

A:Cross-references: GB:X0533; NID:963393; PIDN:CAA29781.1; PID:G295716

A>Note: the authors translated the codon CCG for residue 190 as Gln, CAG for residue 243

C:Genetics:

A:Insertions: 90/1; 129/1; 184/1; 236/1

C:Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II repe

C:Keywords: alternative splicing; duplication; extracellular matrix; glycoprotein; heter

F:1-82/Domain: fibronectin type III repeat homology (fragment) <FN3I>

F:90-172/Domain: fibronectin type III repeat homology <FN3J>

F:167-169/Region: cell attachment (R-G-D) motif

F:184-266/Domain: fibronectin type III repeat homology <FN3K>

Query Match 35.4%; Score 848; DB 2; Length 273;

Best Local Similarity 84.7%; Pred. No. 1.5e-53;

Matches 160; Conservative 15; Mismatches 14; Indels 0; Gaps 0;

QY 89 LDSPTGIDFSIDTANSFTVHWIAPRATITGYRIHHPHFSGRPREDRVPHSRNSTLTN 148

Db 1 LDSPTGLDPSDITANSFTVHWIAPRATITGYRIHHPHGVGRPKEDRVPPSRNSTLTN 60

QY 149 LTPGTYYVVSIVNALNGREESPLIGQOSTVSDVPDRDLEVVAAPTPTSLLSWDAPAVTRY 208

Db 61 LLPGTYYVVSIIAVNGREESVPLVGQTTVSDVPDRDLEVNPTPTSPSTLSISWDAPAVTRY 120

QY 209 YRITVGETGNSVQSFVPGSKSTATISGLKPGVDYTTIVYAVTGRGDSPASSPISIN 268

Db 121 YRITVGETGSSVQSFVPGTMSRATITGLKPGVDYTTIVYAVTGRGDSPASSKPVTVT 180

QY 269 YRTEIDKPS 277

Db 181 YKTEIDTPS 189

RESULT 6

A32398

basic fibroblast growth factor precursor, 22.5K form - human

N:Alternate names: bFGF; fibroblast growth factor 2; prostatic growth factor; prostatrop

N:Contains: basic fibroblast growth factor, 18K form

C:Species: Homo sapiens (man)

C>Date: 31-Jul-1989 #sequence_revision 31-Dec-1993 #text_change 21-Jul-2000

C:Accession: A32398; A26642; B32878; S00297; A54316; B54316; A33624; A25824; B24

R:Pras, H.; Kaghad, M.; Prats, A.C.; Klagebrun, M.; Lelias, J.M.; Liauzun, P.; Chalon,

Proc. Natl. Acad. Sci. U.S.A. 86, 1836-1840, 1989

A>Title: High molecular mass forms of basic fibroblast growth factor are initiated by al

A:Reference number: A32398; MUID:89184522; PMID:2538817

A:Accession: A32398

A:Molecule type: mRNA

A:Residues: 1-210 <PRA>

A:Cross-references: GB:J04513; NID:gl83083; PIDN:AAA52531.1; PID:9459811

R:Shibata, F.; Baird, A.; Florjewska, R.Z.

Growth Factors 4, 277-287, 1991

A>Title: Functional characterization of the human basic fibroblast growth factor gene pr

A:Reference number: A61537; MUID:92110035; PMID:1764264

A:Accession: A61537

A:Molecule type: DNA

A:Residues: 1-114 <SHI>

A>Note: authors translated the codon GGA for residue 47 as Ala

R:Kurokawa, T.; Sasada, R.; Iwane, M.; Igarashi, K.

FEBS Lett. 213, 189-194, 1987

A>Title: Cloning and expression of cDNA encoding human basic fibroblast growth factor.

A:Reference number: A26642; MUID:87162468; PMID:2435575

A:Accession: A26642

A:Molecule type: mRNA

A:Residues: 56-210 <KUR>

A:Cross-references: GB:M27968; NID:gl82562; PIDN:AAA52448.1; PID:gl82563

R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergia, A.; Fiddes, J.C.

Cold Spring Harb. Symp. Quant. Biol. 51, 657-668, 1986

A>Title: Human basic fibroblast growth factor: nucleotide sequence, genomic organization

A:Reference number: A30924; MUID:87217066; PMID:3472745

A:Accession: B32878

A:Molecule type: mRNA

A:Residues: 56-210 <ABR>

A>Note: the authors translated the codon GAA for residue 108 as Gly

R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergia, A.; Friedman, J.; Gospodarowicz, D.; F

EMBO J. 5, 2523-2528, 1986

A>Title: Human basic fibroblast growth factor: nucleotide sequence and genomic organizat

A:Reference number: S00297; MUID:87053817; PMID:3780670

A:Accession: S00297

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-155 <AB2>

A>Note: the authors translated the codon GAA for residue 108 as Gly

R:Shimoyama, Y.; Gotoh, M.; Ino, Y.; Sakamoto, M.; Kato, K.; Hirohashi, S.

Upt. J. Cancer Res. 82, 1263-1270, 1991

A>Title: Characterization of high-molecular-mass forms of basic fibroblast growth factor

rinogenesis.

A:Reference number: A54316; MUID:92091228; PMID:1721615

A:Accession: A54316

A:Molecule type: protein

A:Residues: 'XX', '86-88', 'X', '90-91', 'X', '93-95' <SH3>

A:Experimental source: C-Li21 hepatocellular carcinoma cell line

A>Note: sequence extracted from NCBI backbone (NCBIP:71595)

A:Accession: B54316

A:Molecule type: protein

A:Residues: 'XXX', '19', 'X', '21-29' <SH2>

A>Note: sequence extracted from NCBI backbone (NCBIP:71594)

R:Feige, J.J.; Bradley, J.D.; Fryburg, K.; Farrie, J.; Cousens, L.C.; Barr, P.J.; Baird,

J. Cell Biol. 109, 3105-3114, 1989

A>Title: Differential effects of heparin, fibronectin, and laminin on the phosphorylatic

A:Reference number: A33624; MUID:90078343; PMID:2592418

A:Accession: A33624

A>Status: preliminary

A:Molecule type: protein

A:Residues: 57-210 <PEI>

R:Story, M.T.; Esch, F.; Shimasaki, S.; Sasse, J.; Jacobs, S.C.; Lawson, R.K.

Biochem. Biophys. Res. Commun. 142, 702-709, 1987

A>Title: Amino-terminal sequence of a large form of basic fibroblast growth factor isola

A:Reference number: A25824; MUID:87156686; PMID:2435284

A:Accession: A25824

A:Molecule type: protein

A:Residues: 57-77 <STO>

A:Experimental source: prostate

R:Gimenez-Gallego, G.; Conn, G.; Hatcher, V.B.; Thomas, K.A.

Biochem. Biophys. Res. Commun. 135, 541-548, 1986

A>Title: Human brain-derived acidic and basic fibroblast growth factors: amino terminal

A:Reference number: A30122; MUID:86186784; PMID:3964259

A:Accession: B24243

A:Molecule type: protein

A:Residues: 65-102, 'X', '104-105' <GIM>

A:Experimental source: brain

R:Gautschi, P.; Frater-Schroder, M.; Bohlen, P.

FEBS Lett. 204, 203-207, 1986

A>Title: Partial molecular characterization of endothelial cell mitogens from human brai

A:Reference number: A91364; MUID:86275260; PMID:3732516

A:Accession: B24301

A:Molecule type: protein

A:Residues: 65-88, 'X', '90-98', 'X', '100' <GAU>

R:Sommer, A.; Brewer, M.T.; Thompson, R.C.; Moscatelli, D.; Presta, M.; Rifkin, D.B.

Biochem. Biophys. Res. Commun. 144, 543-550, 1987

A>Title: A form of human basic fibroblast growth factor with an extended amino terminus

A:Reference number: S42242; MUID:87213238; PMID:3579930

A:Accession: S42242
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 54-210 <SOM>
A:Cross-references: EMBL:M17599, NID:g183086; PIDN:AAA52534.1.; PID:g183087
R:Pantoliano, M.W.; Horlick, R.A.; Springer, B.A.; Van Dyk, D.E.; Iobery, T.; Wetmore, D.
Biochemistry 33, 10229-10248, 1994
A>Title: Multivalent ligand-receptor binding interactions in the fibroblast growth factor
A:Reference number: A55784; MUID:94347757; PMID:7520751
A:Accession: B55784
A:Molecule type: protein
A:Residues: 54-71 <PAN>
R:Watson, R.; Anthony, F.; Pickett, M.; Lambden, P.; Masson, G.M.; Thomas, E.J.
Biochem Biophys Res Commun. 187, 1227-1231, 1992
A>Title: Reverse transcription with nested polymerase chain reaction shows expression of
transcripts.

A:Reference number: 152267; MUID:93038590; PMID:1417798
A:Accession: 152267
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 95-182 <RES>
A:Cross-references: GB:S47380; NID:G256535; PIDN:AAD13853.1.; PID:G4261553
A:Experimental source: granulosa cells
R:Patry, V.; Bugler, B.; Amalric, F.; Promet, J.C.; Prats, H.
FEBS Lett. 349, 23-28, 1994
A>Title: Purification and characterization of the 210-amino acid recombinant basic fibro
A:Reference number: S46253; MUID:94320639; PMID:8045296
A:Accession: S46253
A:Molecule type: protein
A:Residues: 39-53;65-88 <PAT>
A>Note: recombinant gene expressed in Escherichia coli
C:Genetics:
A:Gene: GDB:FGF2; FGFB
A:Cross-references: GDB:I19910; OMIM:I34920
A:Map position: 4q25-q27
A:Start codon: CUG
C:Superfamily: fibroblast growth factor
C:Keywords: alternative initiators; angiogenesis; growth factor; heparin binding; mitoge
P:1-210/Product: basic fibroblast growth factor, 22.5K form #status predicted <MA2>
P:65-210/Product: basic fibroblast growth factor, 18K form #status predicted <MAT>
P:82-86/Region: heparin binding #status predicted
P:171-174/Region: heparin binding #status predicted

Query Match 34.4%; Score 824.5; DB 2; Length 210;
Best Local Similarity 88.8%; Pred. No. 5e-52;
Matches 159; Conservative 4; Mismatches 15; Indels 1; Gaps 1;

Qy 254 GRGDSPASSKPISINYTEIDTKPSMAAGSTITPALPEDGGSGAFPPGHFKDKRLYCKN 313
||| : : : : :
Db 33 GRGTAAAPRAAPAARGSRPG-PAGTMAAGSTITPALPEDGGSGAFPPGHFKDKRLYCKN 91
||| : : : : :
Qy 314 GGFFLRTHPDGRVDGVREKSDPHIKLOQAERGWSVIGVCANRYLAMKEDGRLLASKC 373
||| : : : : :
Db 92 GGFFLRTHPDGRVDGVREKSDPHIKLOQAERGWSVIGVCANRYLAMKEDGRLLASKC 151
||| : : : : :
Qy 374 VTDECFFERLESNNYNTYSRKYTSSYVALKRTGYKLGSKTGPQKAILFLPMASAS 432
||| : : : : :
Db 152 VTDECFFERLESNNYNTYSRKYTSSYVALKRTGYKLGSKTGPQKAILFLPMASAKS 210
||| : : : : :

RESULT 7
GBBOS
basic fibroblast growth factor precursor - bovine (fragment)
N:Alternate names: bBGF; kidney-derived growth factor; prostastropin
C:Species: Bos primigenius taurus (cattle)
C>Date: 13-Aug-1986 #sequence revision 02-Jun-1995 #text change 24-Nov-1999
C:Accession: A24663; A32878; A33784; A61550; A61551; A60310; A61094; A01386; A60316; A22
R:Abraham, J.A.; Mergia, A.; Whang, J.L.; Tumolo, A.; Friedman, J.; Hjerrild, K.A.; Gosp
Science 233, 545-548, 1986
A>Title: Nucleotide sequence of a bovine clone encoding the angiogenic protein, basic fi
A:Reference number: A94290; MUID:86261806; PMID:2425435
A:Accession: A24663
A:Molecule type: mRNA

A;Residues: 3-157 <ABR>
A;Cross-references: GB:M13440; NID:g163049; PIDN:AAA30518.1; PID:g163050
A;Experimental source: pituitary gland
R;Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergia, A.; Fiddes, J.C.
Cold Spring Harb. Symp. Quant. Biol. 51, 657-668, 1986
A;Title: Human basic fibroblast growth factor: nucleotide sequence, genomic organization
A;Reference number: A90924; MUID:87217066; PMID:3472745
A;Accession: A32878
A;Molecule type: mRNA
A;Residues: 3-157 <AB2>
R;Milner, P.G.; Li, Y.S.; Hoffman, R.M.; Kodner, C.M.; Siegel, N.R.; Deuel, T.F.
Biochem. Biophys. Res. Commun. 165, 1096-1103, 1989
A;Title: A novel 17 kD heparin-binding growth factor (HBGF-8) in bovine uterus: purification
A;Reference number: A33784; MUID:90121211; PMID:2610682
A;Accession: A33784
A;Molecule type: protein
A;Residues: 1-14 <ML>
A;Note: demonstration of a possible alternative initiator or splice junction
R;Bertolini, J.; Hearn, M.T.W.
Mol. Cell. Endocrinol. 51, 187-199, 1987
A;Title: Isolation, characterization and tissue localisation of an N-terminal-truncated
A;Reference number: A61550; MUID:87247652; PMID:3596000
A;Accession: A61550
A;Molecule type: protein
A;Residues: 16-35 <BER>
R;Ueno, N.; Baird, A.; Esch, F.; Ling, N.; Guillemin, R.
Mol. Cell. Endocrinol. 49, 189-194, 1987
A;Title: Isolation and partial characterization of basic fibroblast growth factor from
A;Reference number: A61551; MUID:87162856; PMID:3556754
A;Accession: A61551
A;Molecule type: protein
A;Residues: 27-35, 'X', 37-41 <UE3>
A;Experimental source: testes
A;Note: this form appears to be identical to the renal form
R;Ueno, N.; Baird, A.; Esch, F.; Shimasaki, S.; Ling, N.; Guillemin, R.
Regul. Pept. 16, 135-145, 1986
A;Title: Purification and partial characterization of a mitogenic factor from bovine li
A;Reference number: A60310; MUID:87119165; PMID:3809608
A;Accession: A60310
A;Molecule type: protein
A;Residues: 23-35, 'X', 37-42 <UEN>
A;Experimental source: liver
R;Ueno, N.; Baird, A.; Esch, F.; Ling, N.; Guillemin, R.
Biochem. Biophys. Res. Commun. 138, 580-588, 1986
A;Title: Isolation of an amino terminal extended form of basic fibroblast growth factor
A;Reference number: A24819; MUID:86295737; PMID:3741423
A;Contents: annotation
R;Gospodarowicz, D.; Baird, A.; Cheng, J.; Lui, G.M.; Esch, F.; Bohlen, P.
Endocrinology 118, 82-90, 1986
A;Title: Isolation of fibroblast growth factor from bovine adrenal gland: physicochemical
A;Reference number: A61094; MUID:86081530; PMID:3940857
A;Accession: A61094
A;Molecule type: protein
A;Residues: 12-25, 27-35, 'X', 37-40 <GOS>
A;Experimental source: adrenal gland
R;Esch, F.; Baird, A.; Ling, N.; Ueno, N.; Hill, F.; Denoroy, L.; Klepper, R.; Gospodarowicz, R.;
Proc. Natl. Acad. Sci. U.S.A. 82, 6507-6511, 1985
A;Title: Primary structure of bovine pituitary basic fibroblast growth factor (FGF) and
A;Reference number: A01386; MUID:86016731; PMID:3863109
A;Accession: A01386
A;Molecule type: protein
A;Residues: 12-157 <ESC>
A;Experimental source: pituitary gland
R;Baird, A.; Esch, F.; Boehlen, P.; Ling, N.; Gospodarowicz, D.
Regul. Pept. 12, 201-213, 1985
A;Title: Isolation and partial characterization of an endothelial cell growth factor fr
A;Reference number: A60316; MUID:86095426; PMID:4081126
A;Accession: A60316
A;Molecule type: protein
A;Residues: 27-35, 'X', 37-43 <BAI>
A;Experimental source: kidney
R;Bohlen, P.; Baird, A.; Esch, F.; Ling, N.; Gospodarowicz, D.

Proc. Natl. Acad. Sci. U.S.A. 81, 5364-5368, 1984
A;Title: Isolation and partial molecular characterization of pituitary fibroblast growth
A;Reference number: A22054; MUID:842298139; PMID:6591194
A;Accession: A22054
A;Molecule type: protein
A;Residues: 12-26 <BOH>
C;Comment: The acidic and basic fibroblast growth factors are the major endothelial-cell
cell types in vitro (although bFGF is 30-100 times more potent than aFGF in stimulating t
C;Superfamily: fibroblast growth factor
C;Comment: This protein binds heparin more strongly than does aFGF.
C;Keywords: fibroblast growth factor
C;Keywords: alternative splicing; angiogenesis; blocked amino end; growth factor; hepari
F;1-157/Product: basic fibroblast growth factor, uterine form #status predicted <MAT1>
F;4-157/Product: basic fibroblast growth factor, pituitary gamma form #status experimen
F;12-157/Product: basic fibroblast growth factor, pituitary alpha form #status experimen
F;16-157/Product: basic fibroblast growth factor, pituitary short form #status predicted
F;23-157/Product: basic fibroblast growth factor, pituitary short form #status predicted
F;27-157/Product: basic fibroblast growth factor, hepatic form #status experimental <MAT
F;29-33,118-121/Region: heparin binding #status predicted
F;4/Modified site: blocked amino end (Ala) (in mature form pituitary gamma) (probably ac

Query Match 33.9%; Score 812; DB 1; Length 157;
Best Local Similarity 97.4%; Pred. No. 2.6e-51;
Matches 152; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 277 SMAAGSITLTPALPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPH 336
Db 2 AMAAGSITLTPALPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPH 61
QY 337 IKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPPERLESNNYNTYRSRK 396
Db 62 IKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPPERLESNNYNTYRSRK 121
QY 397 YTSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
Db 122 YSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 157
RESULT 8
A31674
basic fibroblast growth factor precursor - rat
N;Alternate names: bFGF
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-May-1990 #sequence revision 21-May-1990 #text_change 16-Jul-1999
C;Accession: A31674; S00876; S24309
R;Shimasaki, S.; Emoto, N.; Koba, A.; Mercado, M.; Shibata, F.; Cooke, K.; Baird, A.;
Biochem. Biophys. Res. Commun. 157, 256-263, 1988
A;Title: Complementary DNA cloning and sequencing of rat ovarian basic fibroblast growth
A;Reference number: A31674; MUID:89061721; PMID:3196337
A;Accession: A31674
A;Molecule type: mRNA
A;Residues: 1-154 <SHI>
A;Cross-references: GB:M22427; NID:G204285; PIDN:AAA41210.1; PID:G204286
R;Kurokawa, T.; Seno, M.; Igarashi, K.
Nucleic Acids Res. 16, 5201, 1988
A;Title: Nucleotide sequence of rat basic fibroblast growth factor cDNA.
A;Reference number: S00876; MUID:88462516; PMID:3387229
A;Accession: S00876
A;Molecule type: mRNA
A;Residues: 1-154 <KUR>
A;Cross-references: EMBL:X07285; NID:G56203; PIDN:CAA30265.1; PID:G56204
R;El-Husseini, A.E.D.; Paterson, J.A.; Myal, Y.; Shiu, R.P.C.
Biochim. Biophys. Acta 1131, 314-316, 1992
A;Title: PCR detection of the rat brain basic fibroblast growth factor (bFGF) mRNA conta
A;Reference number: S24309; MUID:92329546; PMID:1378302
A;Accession: S24309
A;Status: preliminary; translation not shown
A;Molecule type: mRNA
A;Residues: 35-154 <ELH>
A;Cross-references: EMBL:X61697; NID:G56143; PIDN:CAA43863.1; PID:G56144
C;Superfamily: fibroblast growth factor
C;Keywords: growth factor
F;1-9/Domain: signal sequence #status predicted <SIG>
F;10-154/Product: basic fibroblast growth factor #status predicted <MAT>

Query Match 33.0%; Score 792.5; DB 2; Length 154;
Best Local Similarity 96.1%; Pred. No. 6.4e-50;
Matches 149; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
QY 278 MAAGSITLTPALPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHI 337
Db 1 MAAGSITSLPALPEDGG-GAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHV 59
QY 338 KLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPPERLESNNYNTYRSKY 397
Db 60 KLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPPERLESNNYNTYRSKY 119
QY 398 TSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
Db 120 SSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 154
RESULT 9
C37360
basic fibroblast growth factor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 17-Apr-1993 #sequence revision 17-Apr-1993 #text_change 16-Jul-1999
C;Accession: C37360
R;Hebert, J.M.; Basilico, C.; Goldfarb, M.; Haub, O.; Martin, G.R.
Dev. Biol. 138, 454-463, 1990
A;Title: Isolation of cDNAs encoding four mouse EGF family members and characterization
A;Reference number: A37360; MUID:90201563; PMID:2318343
A;Accession: C37360
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-154 <HEB>
A;Cross-references: GB:M30644; NID:G193296; PIDN:AAA37621.1; PID:G309239
C;Superfamily: fibroblast growth factor
Query Match 32.4%; Score 777.5; DB 2; Length 154;
Best Local Similarity 94.2%; Pred. No. 7.7e-49;
Matches 146; Conservative 5; Mismatches 3; Indels 1; Gaps 1;
QY 278 MAAGSITLTPALPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHI 337
Db 1 MAAGSITSLPALPEDGGA-APPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHV 59
QY 338 KLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPPERLESNNYNTYRSKY 397
Db 60 KLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPPERLESNNYNTYRSKY 119
QY 398 TSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
Db 120 SSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 154
RESULT 10
S71465
fibronectin - chicken (fragment)
C;Species: Gallus gallus (Chicken)
C;Date: 12-Feb-1998 #sequence revision 13-Mar-1998 #text_change 21-Jan-2000
C;Accession: S71465
R;Gehris, A.L.; Brandli, D.W.; Lewis, S.D.; Bennett, V.D.
Biochim. Biophys. Acta 1311, 5-12, 1996
A;Title: The exon encoding the fibronectin type III-9 repeat is constitutively included
A;Reference number: S71465; MUID:96183658; PMID:8603103
A;Accession: S71465
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-189 <GEH>
A;Cross-references: EMBL:U20386; NID:G1323735; PIDN:AAB01062.1; PID:G1323736
A;Note: the authors translated the codon AAG for residue 50 as Asn and AAG for residue 1
C;Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II repe
F;100-183/Domain: fibronectin type III repeat homology <3FR>
Query Match 32.0%; Score 768; DB 2; Length 189;
Best Local Similarity 83.1%; Pred. No. 4.9e-48;


```
Query Match      30.8%; Score 738; DB 2; Length 137;
Best Local Similarity 99.3%; Pred. No. 4.5e-46;
Matches 136; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 287 PALPEDGGSGAPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEER 346
Db 1 PALPEDGGSGAPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEER 60

QY 347 GVVSIGKVCANRYLAMKEDGRLLASKCVTDEGFFERLESNNYNTYRSRKYTSWYVALKR 406
Db 61 GVVSIGKVCANRYLAMKEDGRLLASKCVTDEGFFERLESNNYNTYRSRKYTSWYVALKR 120

QY 407 TGQYKLGSKTGPQKAI 423
Db 121 TGQYKLGSKTGPQKAI 137

RESULT 15
A40117
basic fibroblast growth factor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A40117; A29618
R:Kimelman, D.; Abraham, J.A.; Haaparanta, T.; Palisi, T.M.; Kirschner, M.W.
Science 242, 1053-1056, 1988
A:Title: The presence of fibroblast growth factor in the frog egg: its role as a natural
A:Reference number: A40117; MUID:89058621; PMID:3194757
A:Accession: A40117
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-155 <KIM>
A:Cross-references: GB:M18067; NID:G214177; PIDN:AAA49726.1; PID:G214178; GB:M21092
R:Kimelman, D.; Kirschner, M.
Cell 51, 869-877, 1987
A:Title: Synergistic induction of mesoderm by FGF and TGF-beta and the identification of
A:Reference number: A29618; MUID:88052890; PMID:3479265
A:Accession: A29618
A:Molecule type: mRNA
A:Residues: 95-110,112-155 <KI2>
C:Superfamily: fibroblast growth factor
C:Keywords: growth factor

Query Match      28.4%; Score 681; DB 1; Length 155;
Best Local Similarity 83.2%; Pred. No. 6.6e-42;
Matches 129; Conservative 9; Mismatches 17; Indels 0; Gaps 0;

QY 278 MAAGSITTLPALPEDGGSGAPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHI 337
Db 1 MAAGSITTLPTSESDGNTFPFGSFKDPKRLYCKNGGFFLRINS DGRVDGSRDKSDSHI 60

QY 338 KLQQAEEERGVSIGKVCANRYLAMKEDGRLLASKCVTDEGFFERLESNNYNTYRSKY 397
Db 61 KLQQAEEERGVSIGKITANRYLAMKEDGRLLSKRCITDEGFFERLEANNYNTYRSKY 120

QY 398 TSWYVALKRTGQYKLGSKTGPQKAILFLPMSAAS 432
Db 121 TSWYVALKRTGQYKLGSGSTGPQKAILFLPMSAKS 155

Search completed: May 3, 2004, 13:11:48
Job time : 21.9005 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:01:53 ; Search time 14.9253 Seconds
(without alignments)
1594.339 Million cell updates/sec

Title: US-09-775-964-5
Perfect score: 2398
Sequence: 1 PTDLRFTNIGPDMRTWAP.....LVTLPHPNLHGPEILDVPST 457

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1474	61.5	2386	1	FINC HUMAN
2	1396	58.2	2265	1	FINC BOVIN
3	1345	56.1	2477	1	FINC RAT
4	1340	55.9	2477	1	FINC MOUSE
5	1210.5	50.5	1256	1	FINC CHICK
6	1051.5	43.8	1328	1	FINC PLEWA
7	1020.5	42.6	2481	1	FINC XENLA
8	820	34.2	155	1	FGF2_HUMAN
9	811	33.8	155	1	FGF2_BOVIN
10	805	33.6	155	1	FGF2 SHEEP
11	792.5	33.0	154	1	FGF2 RAT
12	777.5	32.4	154	1	FGF2 MONDO
13	754.5	31.5	156	1	FGF2 CHICK
14	753	31.4	158	1	FGF2_HUMAN
15	738	30.8	137	1	FGF2_RABIT
16	681	28.4	155	1	FGF2 XENLA
17	425.5	17.7	155	1	FGF1 MESAU
18	417.5	17.4	155	1	FGF1 CHICK
19	416.5	17.4	155	1	FGF1_HUMAN
20	411.5	17.2	155	1	FGF1_MOUSE
21	403.5	16.8	152	1	FGF1_PIG
22	400.5	16.7	155	1	FGF1_BOVIN
23	354.5	14.8	3063	1	CALC1 HUMAN
24	354	14.8	1560	1	TENN MOUSE
25	345.5	14.4	3119	1	CALC1_MOUSE
26	339.5	14.2	639	1	CALC1_RABIT
27	338.5	14.1	1746	1	TENA_PIG
28	334.5	13.9	2201	1	TENA HUMAN
29	333.5	13.9	929	1	CALC1_NOTVI
30	331	13.8	1808	1	TENA_CHICK
31	325	13.6	1294	1	TENN_HUMAN
32	309	12.9	522	1	FINC CANFA
33	309	12.9	3124	1	CALC1_CHICK

RESULT 1

ID	FINC_HUMAN	STANDARD;	PRT;	2386 AA.
AC	P02751; O95609; Q95610; Q14312; Q14326; Q86T27; Q8R1V8;			
AC	Q96KP7; Q96KP8; Q96KP9; Q9H1B8; Q9HAP3; Q9UMK2;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Fibronectin precursor (FN) (Cold-insoluble globulin) (CIG).			
GN	FN1 OR FN.			
OS	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
OC	NCBI_TaxId=9606;			
OX	[1]			
RN	SEQUENCE FROM N.A. (ISOFORM 2).			
RX	MEDLINE=21600194; PubMed=11737888;			
RA	Schor S.L., Schor A.M.;			
RT	"phenotypic and genetic alterations in mammary stroma: implications for tumour progression.";			
RL	Breast Cancer Res. 3:373-379(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORMS 3; 7 AND 10).			
RC	TISSUE=Cervix;			
RA	Ansorge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,			
RA	Mewes H.-W., Weill B., Amid C., Osanger A., Fobo G., Han M.,			
RA	Wiemann S.;			
RN	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.			
RP	[3]			
RP	SEQUENCE OF 1-38 FROM N.A.			
RP	MEDLINE=87030890; PubMed=3770189;			
RA	Gutman A., Yamada K.M., Kornblitt A.R.;			
RT	"Human fibronectin is synthesized as a pre-propolypeptide.";			
RL	FEBS Lett. 207:145-148(1986).			
RN	[4]			
RP	SEQUENCE OF 1-49 FROM N.A.			
RP	MEDLINE=87175578; PubMed=3031656;			
RA	Dean D.C., Bowls C.L., Bourgeois S.;			
RT	"Cloning and analysis of the promoter region of the human fibronectin gene.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 84:1876-1880(1987).			
RN	[5]			
RP	SEQUENCE OF 28-2386 FROM N.A. (ISOFORM 3).			
RX	MEDLINE=85284965; PubMed=2992939;			
RA	Kornblitt A.R., Umezawa K., Vibe-Pedersen K., Baralle F.E.;			
RT	"Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene.";			
RL	EMBO J. 4:1755-1759(1985).			
RN	[6]			
RP	SEQUENCE OF 103-481 AND 2228-2386 FROM N.A. (ISOFORMS 1; 3; 8 AND 9).			
RC	TISSUE=Periphereal blood T-cell, and Umbilical vein endothelial cells;			
RA	Godfrey H.P., Ebrahim A.A.;			
RL	Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.			
RN	[7]			
RP	SEQUENCE OF 973-2386 FROM N.A. (ISOFORM 3).			
RX	MEDLINE=84272258; PubMed=6462919;			

34	307	12.8	522	1	FINC HORSE
35	290.5	12.1	1888	1	CALC_CHICK
36	283	11.8	4289	1	TENN_HUMAN
37	278	11.6	1329	1	KF10_HUMAN
38	272	11.3	2944	1	CA17_HUMAN
39	270.5	11.3	206	1	FGF4_HUMAN
40	266	11.1	194	1	FGF4_CHICK
41	264	11.0	220	1	FGF3_CHICK
42	258.5	10.8	206	1	FGF4_BOVIN
43	257.5	10.7	266	1	FGF5 RAT
44	256.5	10.7	264	1	FGF5_MOUSE
45	255.5	10.7	256	1	FGF3_BRARE

ALIGNMENTS

RA Kornblith A.R., Vibe-Pedersen K., Baralle F.E.;
RT "Human fibronectin: cell specific alternative mRNA splicing generates
RT polypeptide chains differing in the number of internal repeats.";
RL Nucleic Acids Res. 12:5853-5868(1984).
RN [8]
RP SEQUENCE OF 1232-1782 FROM N.A. (ISOFORM 7).
RX MEDLINE=88233940; PubMed=3375063;
RA Paoletti G., Henschcliff C., Sebastio G., Baralle F.E.;
RT "Sequence analysis and in vivo expression show that alternative
RT splicing of ED-B and ED-A regions of the human fibronectin gene are
RT independent events.";
RL Nucleic Acids Res. 16:3545-3557(1988).
RN [9]
RP SEQUENCE OF 1257-1365 FROM N.A. (ISOFORM 11).
RX MEDLINE=88041070; PubMed=3478690;
RA Gutman A., Kornblith A.R.;
RT "Identification of a third region of cell-specific alternative
RT splicing in human fibronectin mRNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:7179-7182(1987).
RN [10]
RP SEQUENCE OF 1441-1548.
RX MEDLINE=82265604; PubMed=7050098;
RA Pierchbacher M.D., Ruoslahti E., Sundelin J., Lind P., Peterson P.A.;
RT "The cell attachment domain of fibronectin. Determination of the
RT primary structure.";
RL J. Biol. Chem. 257:9593-9597(1982).
RN [11]
RP SEQUENCE OF 1448-1540 FROM N.A.
RX MEDLINE=83290929; PubMed=6688418;
RA Oldberg A., Linney E., Ruoslahti E.;
RT "Molecular cloning and nucleotide sequence of a cDNA clone coding for
RT the cell attachment domain in human fibronectin.";
RL J. Biol. Chem. 258:10193-10196(1983).
RN [12]
RP SEQUENCE OF 1448-1540 FROM N.A.
RX MEDLINE=86111901; PubMed=3003095;
RA Oldberg A., Ruoslahti E.;
RT "Evolution of the fibronectin gene. Exon structure of cell attachment
RT domain.";
RL J. Biol. Chem. 261:2113-2116(1986).
RN [13]
RP SEQUENCE OF 1594-2386 FROM N.A. (ISOFORM 1).
RX MEDLINE=85280409; PubMed=2992573;
RA Bernard M.P., Kolbe M., Weil D., Chu M.-L.;
RT "Human cellular fibronectin: comparison of the carboxyl-terminal
RT portion with rat identifies primary structural domains separated by
RT hypervariable regions.";
RL Biochemistry 24:2698-2704(1985).
RN [14]
RP SEQUENCE OF 1712-1739 FROM N.A.
RX MEDLINE=87026578; PubMed=3021206;
RA Sakiguchi K., Klos A.M., Kurachi K., Yoshitake S., Hakomori S.;
RT "Human liver fibronectin complementary DNAs: identification of two
RT different messenger RNAs possibly encoding the alpha and beta
RT subunits of plasma fibronectin.";
RL Biochemistry 25:4936-4941(1986).
RN [15]
RP SEQUENCE OF 1788-2386 FROM N.A. (ISOFORMS 4; 5 AND 6).
RX TISSUE=cartilage; PubMed=12127832;
RA Parker A.E., Boutell J., Carr A., Maciewicz R.A.;
RT "Novel cartilage-specific splice variants of fibronectin.";
RL Osteoarthritis Cartilage 10:528-534(2002).
RN [16]
RP SEQUENCE OF 32-290.
RX MEDLINE=84032463; PubMed=6630202;
RA Garcia-Pardo A., Pearlstein E., Frangione B.;
RT "Primary structure of human plasma fibronectin. The 29,000-dalton
RT NH2-terminal domain.";
RL J. Biol. Chem. 258:12670-12674(1983).
RN [17]
RP SEQUENCE OF 309-608, AND COLLAGEN-BINDING.
RX MEDLINE=87080265; PubMed=3024962;
RA Owens R.J., Baralle F.E.;
RT "Mapping the collagen-binding site of human fibronectin by expression
RT in Escherichia coli.";
RL EMBO J. 5:2825-2830(1986).
RN [18]
RP SULFATION.
RX MEDLINE=86042625; PubMed=2414772;
RA Liu M.C., Yu S., Sy J., Redman C.M., Lipmann F.;
RT "Tyrosine sulfation of proteins from the human hepatoma cell line
RT HepG2.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7160-7164(1985).
RN [19]
RP O-GLYCOSYLATION OF THR-2064.
RX MEDLINE=91190085; PubMed=2012601;
RA Tressell T., McCarthy J.B., Calaycay J., Lee T.D., Legesse K.,
RT Shively J.E., Pande H.;
RT "Human plasma fibronectin. Demonstration of structural differences
RT between the A- and B-chains in the III CS region.";
RL Biochem. J. 274:731-738(1991).
RN [20]
RP FBN1-BINDING SITE.
RX MEDLINE=93015879; PubMed=1400330;
RA Baibona K., Tran H., Godyna S., Ingham K.C., Strickland D.K.,
RT Argaves W.S.;
RT "Fibulin binds to itself and to the carboxyl-terminal heparin-binding
RT region of fibronectin.";
RL J. Biol. Chem. 267:20120-20125(1992).
RN [21]
RP CHARACTERIZATION OF FIBRIN-BINDING SITE 1.
RX MEDLINE=95081153; PubMed=7989369;
RA Rostagno A., Williams M.J., Baron M., Campbell I.D., Gold L.I.;
RT "Further characterization of the NH2-terminal fibrin-binding site on
RT fibronectin.";
RL J. Biol. Chem. 269:31938-31945(1994).
RN [22]
RP STRUCTURE BY NMR OF 1447-1540.
RX MEDLINE=92162710; PubMed=1311202;
RA Baron M., Main A.L., Driscoll P.C., Mardon H.J., Boyd J.,
RT Campbell I.D.;
RT "1H NMR assignment and secondary structure of the cell adhesion type
RT III module of fibronectin.";
RL Biochemistry 31:2068-2073(1992).
RN [23]
RP STRUCTURE BY NMR OF 1447-1540.
RX MEDLINE=93046665; PubMed=1423622;
RA Main A.L., Harvey T.S., Baron M., Boyd J., Campbell I.D.;
RT "The three-dimensional structure of the tenth type III module of
RT fibronectin: an insight into RGD-mediated interactions.";
RL Cell 71:671-678(1992).
RN [24]
RP STRUCTURE BY NMR OF 182-275.
RX MEDLINE=94141923; PubMed=8308892;
RA Williams M.J., Phan I., Harvey T.S., Rostagno A., Gold L.I.,
RA Campbell I.D.;
RT "Solution structure of a pair of fibronectin type 1 modules with
RT fibrin binding activity.";
RL J. Mol. Biol. 235:1302-1311(1994).
RN [25]
RP STRUCTURE BY NMR OF 32-92.
RX MEDLINE=96069779; PubMed=7583666;
RA Potts J.R., Phan I., Williams M.J., Campbell I.D.;
RT "High-resolution structural studies of the factor XIIIa crosslinking
RT site and the first type 1 module of fibronectin.";
RL Nat. Struct. Biol. 2:946-950(1995).
RN [26]
RP STRUCTURE BY NMR OF 406-464.
RX MEDLINE=98179558; PubMed=9514732;
RA Sticht H., Pickford A.R., Potts J.R., Campbell I.D.;
RT "Solution structure of the glycosylated second type 2 module of
RT fibronectin.";
RL J. Mol. Biol. 276:177-187(1998).
RN [27]
RP STRUCTURE BY NMR OF EXTRA ED-B DOMAIN FROM ISOFORM 7.

```
Query Match 61.5%, Score 1474, DB 1, Length 2386;
Best Local Similarity 46.8%, Pred. No. 5.4e-96;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY 1 PTDLRFNIGDPTWRTVWAPPSPSIDLNFVLRVSPVKNEDVAELSTSPSDNAVLTNLL 60
DB 1270 PTDLRFNIGDPTWRTVWAPPSPSIDLNFVLRVSPVKNEDVAELSTSPSDNAVLTNLL 1329

QY 61 PGTEYVSVSVYEQHSTPLRGKQKTGLDPSGTIDFSDITANSFTVHMTAPRATITGYR 120
DB 1330 PGTEYVSVSVYEQHSTPLRGKQKTGLDPSGTIDFSDITANSFTVHMTAPRATITGYR 1389

QY 121 IRHHPFSGRPREDRVPHSRNSTLTNLTGTEYVSVSVYEQHSTPLRGKQKTGLDPSGTIDFSDITANSFTVHMTAPRATITGYR 180
DB 1390 IRHHPFSGRPREDRVPHSRNSTLTNLTGTEYVSVSVYEQHSTPLRGKQKTGLDPSGTIDFSDITANSFTVHMTAPRATITGYR 1449

QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITTYGETGNSPVQSEFTVPGSKSTATISGLK 240
DB 1450 VPRDLEVAATPTSLISWDAPAVTVRYRITTYGETGNSPVQSEFTVPGSKSTATISGLK 1509

QY 241 PGVDYITTVAVTGRGDSPPASSKPSISINRYTEIDKPS-----M 278
DB 1510 PGVDYITTVAVTGRGDSPPASSKPSISINRYTEIDKPS-----M 278

QY 279 AAGSITTL-----PDGRV 326
DB 1570 TGYRVITTPKNGPPTKTAGDQTEMTIEGLQPTVEYVSVVYVQAQNPSSQPLVQTAV 1629

QY 288 -----ALPDGSGGAFPPGHPKPKRL 309
DB 1630 TNIDRPKGLAFTVDVDSIKIAMESPGQVSRVRYTVSPEDGHELFAPDGEEDTAEL 1689

QY 310 YKNGG-----PDGRV 326
DB 1690 QGLRPGSEYTVSVVALHDDMESQPLICTQSTALPAPDLKFTQVTPSLSAQWTPNVQL 1749

QY 327 DGVR-----EKSDFHKLQAEERGVSIGKVCANRY-----LAWKEDGRLLASK 372
DB 1750 TGYRVVTPKREKTPGKMEINLAPDSSVSVVGLMVATKYEVSVVALKD---TLTSRPAQ 1806

QY 373 -----CVTDEC-----FFERLESNNYNTYRS-----R 395
DB 1807 VTTLENVSPRARVTDATETITISWRTKTETITGQVDAVPANGQTPQIKPDVR 1866

QY 396 KYT-----SWYVA 403
DB 1867 SYITGLQPGTDYKIVLYTLNDNARSPVVIDASTAIDAPSNLRFATTPNSLLVSWQPP 1926

QY 404 LKRTGQY-----KLGSK-----TG--PGQKAILFL-----PM- 428
DB 1927 RARITGYIIEKYPGSPPREWPRPRPGVTETATITGLEPGTEYTVIVIALKNNQKSEPLI 1986

QY 429 -SAASDELPOVLTPHNLHGPILDPVST 457
DB 1987 GRKKTDELPOVLTPHNLHGPILDPVST 2016
```

RESULT 2

```
ID FINE_BOVIN STANDARD; PRT; 2265 AA.
AC P07589;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fibronectin (FN).
GN FN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
```

```
RP SEQUENCE.
RX MEDLINE=87054047; PubMed=3780752;
RA Skorstengaard K., Jensen M.S., Sahl P., Petersen T.E., Magnusson S.;
RT "Complete primary structure of bovine plasma fibronectin.";
RN Eur. J. Biochem. 161:441-453(1986).
RP PARTIAL SEQUENCE.
RX MEDLINE=83117805; PubMed=6218503;
RA Petersen T.E., Thorgersen H.C., Skorstengaard K., Vibe-Pedersen K.,
RA Sahl P., Sottrup-Jensen L., Magnusson S.;
RT "Partial primary structure of bovine plasma fibronectin: three types
of internal homology.";
RN Proc. Natl. Acad. Sci. U.S.A. 80:137-141(1983).
RP SEQUENCE OF 2170-2265 FROM N.A.
RX MEDLINE=83221567; PubMed=6304699;
RA Kornblitt A.R., Vibe-Pedersen K., Baralle F.E.;
RT "Isolation and characterization of cDNA clones for human and bovine
fibronectin.";
RN Proc. Natl. Acad. Sci. U.S.A. 80:3218-3222(1983).
CC -!- FUNCTION: Fibronectins bind cell surfaces and various compounds
including collagen, fibrin, heparin, DNA, and actin. Fibronectins
are involved in cell adhesion, cell motility, opsonization, wound
healing, and maintenance of cell shape.
CC -!- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED
VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;
CC TO A LESSER EXTENT HOMODIMERS.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. Each of the "extra
domain" and the connecting strand 3 are present in some forms of
fibronectin and absent in others;
CC Name=1;
CC IsoId=P07589-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted
by hepatocytes. Cellular FN (dimeric or cross-linked multimeric
forms), made by fibroblasts, epithelial and other cell types, is
deposited as fibrils in the extracellular matrix.
CC -!- PTM: Sulfated (By similarity).
CC -!- SIMILARITY: Contains 12 fibronectin type I domains.
CC -!- SIMILARITY: Contains 2 fibronectin type II domains.
CC -!- SIMILARITY: Contains 15 fibronectin type III domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: K00800; AAA30521.2; --
CC FIR; A26452; FNBO.
CC HSSP; P02751; 2FN2.
CC InterPro: IPR006209; EGF like.
CC InterPro: IPR000083; Fibnctnl.
CC InterPro: IPR008957; FN III-like.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR000562; FN_Type_II.
CC InterPro: IPR003962; FNIII_subd.
CC Pfam; PF00039; fn1; 2.
CC Pfam; PF00040; fn2; 2.
CC Pfam; PF00041; fn3; 15.
CC PRINTS; PR00013; FNTYPEII.
CC PRODOM; PR00014; FNTYPEIII.
CC SMART; SM00058; FN1; 12.
CC SMART; SM00059; FN2; 2.
CC SMART; SM00060; FN3; 14.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS00023; FIBRONECTIN_2; 2.
CC PROSITE; PS01253; FIBRONECTIN_1; 12.
```


GN FN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer; TISSUE=Liver;
RX MEDLINE=88054951; PubMed=2445560;
RA Schwarzbaumer J.E., Patel R.S., Fonda D., Hynes R.O.;
RT "Multiple sites of alternative splicing of the rat fibronectin gene
transcript.";
RL EMBO J. 6:2573-2580(1987).
RN [2]
RP SEQUENCE OF 1-139 AND 2382-2477 FROM N.A.
RC STRAIN=Fischer; TISSUE=Liver;
RX MEDLINE=88054950; PubMed=3119323;
RA Patel R.S., Odermatt E., Schwarzbaumer J.E., Hynes R.O.;
RT "Organization of the fibronectin gene provides evidence for exon
shuffling during evolution.";
RL EMBO J. 6:2565-2572(1987).
RN [3]
RP SEQUENCE OF 1586-2477 FROM N.A.
RX MEDLINE=84082067; PubMed=6311787;
RA Schwarzbaumer J.E., Tamkun J.W., Lemischka I.R., Hynes R.O.;
RT "Three different fibronectin mRNAs arise by alternative splicing
within the coding region.";
RL Cell 35:421-431(1983).
CC -1- FUNCTION: Fibronectins bind cell surfaces and various compounds
including collagen, fibrin, heparin, DNA, and actin. Fibronectins
are involved in cell adhesion, cell motility, opsonization, wound
healing, and maintenance of cell shape.
CC -1- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED
VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;
CC TO A LESSER EXTENT HOMODIMERS.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Each of the "extra domain" and the connecting strand 3
CC are present in some forms of fibronectin and absent in others;
CC Name=1;
CC IsoId=P04937-1; Sequence=Displayed;
CC Name=2; Synonyms=FNIII-13-16ss;
CC IsoId=P04937-2; Sequence=VSP_003258;
CC Name=3; Synonyms=Lambda-RLP4-5;
CC IsoId=P04937-3; Sequence=VSP_003259;
CC Name=4; Synonyms=Lambda-RLP6;
CC IsoId=P04937-4; Sequence=VSP_003260;
CC -1- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted
by hepatocytes. Cellular FN (dimeric or cross-linked multimeric
forms), made by fibroblasts, epithelial and other cell types, is
deposited as fibrils in the extracellular matrix.
CC -1- PTM: Sulfated (by similarity).
CC -1- SIMILARITY: Contains 12 fibronectin type I domains.
CC -1- SIMILARITY: Contains 2 fibronectin type II domains.
CC -1- SIMILARITY: Contains 17 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15906; CAA34020.1; --
DR EMBL; L29191; AAA41166.1; --
DR EMBL; L00191; AAA41166.1; JOINED.
DR EMBL; L29191; AAA41167.1; --
DR EMBL; L00191; AAA41167.1; JOINED.
DR EMBL; L29191; AAA41168.1; --
DR EMBL; L00191; AAA41168.1; JOINED.
DR EMBL; X05831; CAA29278.1; --

DR EMBL; X05832; CAA29279.1; --
DR EMBL; X05833; CAA29280.1; --
DR EMBL; X05834; CAA29281.1; --
DR PIR; S14428; S14428.
DR HSP; P02751; IFBR.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000083; Fibnctni.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR000562; FN Type II.
DR InterPro; IPR003962; FNIII_subd.
DR Pfam; PF00039; fn1; 12.
DR Pfam; PF00040; fn2; 2.
DR Pfam; PF00041; fn3; 17.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00014; FNTYPEIII.
DR ProDom; PD000995; FN Type II; 2.
DR SMART; SM00058; FN1; 12.
DR SMART; SM00059; FN2; 2.
DR SMART; SM00060; FN3; 13.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS00023; FIBRONECTIN 2; 2.
DR PROSITE; PS01253; FIBRONECTIN 1; 12.
KW Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;
KW Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal.
FT SIGNAL 1 32
FT CHAIN 33 2477 FIBRONECTIN.
FT DOMAIN 53 273 FIBRIN- AND HEPARIN-BINDING 1.
FT DOMAIN 308 608 COLLAGEN-BINDING.
FT DNA_BIND 906 1171
FT DOMAIN 1357 1630 CELL-ATTACHMENT.
FT DOMAIN 1811 2081 HEPARIN-BINDING 2.
FT DOMAIN 2296 2427 FIBRIN-BINDING 2.
FT DOMAIN 51 91 FIBRONECTIN TYPE-I 1.
FT DOMAIN 96 139 FIBRONECTIN TYPE-I 2.
FT DOMAIN 140 183 FIBRONECTIN TYPE-I 3.
FT DOMAIN 185 229 FIBRONECTIN TYPE-I 4.
FT DOMAIN 230 274 FIBRONECTIN TYPE-I 5.
FT DOMAIN 306 345 FIBRONECTIN TYPE-I 6.
FT DOMAIN 345 404 FIBRONECTIN TYPE-II 1.
FT DOMAIN 405 469 FIBRONECTIN TYPE-II 2.
FT DOMAIN 468 511 FIBRONECTIN TYPE-I 7.
FT DOMAIN 516 558 FIBRONECTIN TYPE-I 8.
FT DOMAIN 559 602 FIBRONECTIN TYPE-I 9.
FT DOMAIN 609 706 FIBRONECTIN TYPE-III 1.
FT DOMAIN 707 808 FIBRONECTIN TYPE-III 2.
FT DOMAIN 809 903 FIBRONECTIN TYPE-III 3.
FT DOMAIN 904 994 FIBRONECTIN TYPE-III 4.
FT DOMAIN 995 1084 FIBRONECTIN TYPE-III 5.
FT DOMAIN 1085 1172 FIBRONECTIN TYPE-III 6.
FT DOMAIN 1173 1264 FIBRONECTIN TYPE-III 7.
FT DOMAIN 1265 1355 FIBRONECTIN TYPE-III 8 (EXTRA DOMAIN 1).
FT DOMAIN 1356 1446 FIBRONECTIN TYPE-III 9.
FT DOMAIN 1447 1536 FIBRONECTIN TYPE-III 10.
FT DOMAIN 1537 1630 FIBRONECTIN TYPE-III 11.
FT DOMAIN 1631 1720 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1721 1810 FIBRONECTIN TYPE-III 13 (EXTRA DOMAIN 2).
FT DOMAIN 1811 1902 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1903 1991 FIBRONECTIN TYPE-III 15.
FT DOMAIN 1992 2081 FIBRONECTIN TYPE-III 16.
FT DOMAIN 2082 2201 CONNECTING STRAND 3 (CS-3) (V REGION).
FT DOMAIN 2202 2283 FIBRONECTIN TYPE-III 17.
FT DOMAIN 2294 2338 FIBRONECTIN TYPE-I 10.
FT DOMAIN 2339 2381 FIBRONECTIN TYPE-I 11.
FT DOMAIN 2382 2426 FIBRONECTIN TYPE-I 12.
FT SITE 1614 1616 CELL ATTACHMENT SITE.
FT SITE 2181 2183 CELL ATTACHMENT SITE.
FT DISULFID 53 79 BY SIMILARITY.
FT DISULFID 77 88 BY SIMILARITY.
FT DISULFID 98 126 BY SIMILARITY.
FT DISULFID 124 136 BY SIMILARITY.
FT DISULFID 142 170 BY SIMILARITY.
FT DISULFID 168 180 BY SIMILARITY.

```
FT DISULFID 187 216 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 232 261 BY SIMILARITY.
FT DISULFID 259 271 BY SIMILARITY.
FT DISULFID 308 335 BY SIMILARITY.
FT DISULFID 333 342 BY SIMILARITY.
FT DISULFID 360 386 BY SIMILARITY.
FT DISULFID 374 401 BY SIMILARITY.
FT DISULFID 420 446 BY SIMILARITY.
FT DISULFID 434 461 BY SIMILARITY.
FT DISULFID 470 498 BY SIMILARITY.
FT DISULFID 496 508 BY SIMILARITY.
FT DISULFID 518 545 BY SIMILARITY.
FT DISULFID 543 555 BY SIMILARITY.
FT DISULFID 561 589 BY SIMILARITY.
FT DISULFID 587 599 BY SIMILARITY.
FT DISULFID 2296 2325 BY SIMILARITY.
FT DISULFID 2323 2335 BY SIMILARITY.
FT DISULFID 2341 2368 BY SIMILARITY.
FT DISULFID 2366 2378 BY SIMILARITY.
FT DISULFID 2385 2409 BY SIMILARITY.
FT DISULFID 2407 2423 BY SIMILARITY.
FT DISULFID 2458 2458 INTERCHAIN (WITH C-2452) .
FT DISULFID 2462 2462 INTERCHAIN (WITH C-2458) .
FT MOD_RES 875 875 SULFATION (POTENTIAL) .
FT MOD_RES 880 880 SULFATION (POTENTIAL) .
FT MOD_RES 2392 2392 SULFATION (POTENTIAL) .
FT MOD_RES 2475 2475 PHOSPHORYLATION (BY SIMILARITY) .
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (BY SIMILARITY) .
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (BY SIMILARITY) .
FT CARBOHYD 542 542 N-LINKED (GLCNAC. .) (BY SIMILARITY) .
FT CARBOHYD 876 876 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1006 1006 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1243 1243 N-LINKED (GLCNAC. .) (BY SIMILARITY) .
FT CARBOHYD 2198 2198 N-LINKED (GLCNAC. .) (BY SIMILARITY) .
FT CARBOHYD 2154 2154 O-LINKED (GALNAC. .) (BY SIMILARITY) .
FT VARSPLIC 1720 1809 Missing (in isoform 2) .
FT VARSPLIC 2082 2106 /FTid=VSP 003258.
FT VARSPLIC 2082 2106 Missing (in isoform 3) .
FT VARSPLIC 2082 2200 /FTid=VSP 003259.
FT CONFLICT 2318 2318 Missing (in isoform 4) .
FT SEQUENCE 2477 AA; 272510 MW; B4391A472ECDEB5 CRC64;

Query Match 56.1%; Score 1345; DB 1; Length 2477;
Best Local Similarity 42.4%; Pred. No. 7.6e-87;
Matches 318; Conservative 49; Mismatches 87; Indels 296; Gaps 18;

QY 1 PTDLRFNTNIGPDMRVMTWAPPSPIDLTNFLVYSPVKNEDVAELISPSDNAVLITNLL 60
DB 1360 PTDLRFNTNIGPDMRVMTWAPPSPIDLTNFLVYSPVKNEDVAELISPSDNAVLITNLL 1419

QY 61 PGTEYVSVSSVYEQHESPLRGRQKTGLDPSPTGIDFSDITANSFTVHWIAPRATITGYR 120
DB 1420 PGTEYLVSVSSVYEQHESPLRGRQKTGLDPSPTGIDFSDITANSFTVHWIAPRATITGYI 1479

QY 121 IRHPHFSGRPREDRVPHSRNSITLNTLPGTEYVSVIVALNGREESPLILGQOSTVSD 180
DB 1480 IRHAEHSAGRPQDRVPPSRNSITLNTLPGTEYIVTIIAVNGREESPLILGQOSTVSD 1539

QY 181 VPRLDLEVAATPTSLISWDAPAVTVYRYRITVGETGNSPVQEFVPGSKSTATISGLK 240
DB 1540 VPRLDLEVIATPTSLISWEPAPVAVRYRITYGETGNSPVQEFVPGSKSTATINNIK 1599

QY 241 PGVDYITVYAVTGRGDSPASSKPIISYRTEIDKPS-----M 278
DB 1600 PGADYITVYAVTGRGDSPASSKPIISYRTEIDKPSQMQVTDQDQNSISVRWLPSPV 1659

QY 279 AGSITLTP----- 287
DB 1660 TGYRVTTAPKNGLGPTKSTQVSPDQTEMTIEGLQPTVYVSVYAQRNGESQPLVQTAV 1719
```

```
QY 288 -----ALPEDGGSGAFP-PGHFKDPKRL 309
DB 1720 TNIDRPKGLAFTDVEDSIKIAWESPOQVSKRYVTYSSPEDGIHELFPDGDDEDTAEL 1779

QY 310 Y-----CKNGGF-----FLRIH-----PDGRV 326
DB 1780 HGLRPGSEYTVSVVALHGGMESQPLIGVQSTAIAPTNLKFTQVSTTLTAQWTAPSVKL 1839

QY 327 DGVR-----EKSDPHIKLQLOAEBGVVSIKVCANRY-----LAMKEDGRLLASK----- 372
DB 1840 TGYRVVTPKKTGPMKEINLSPDSTSVISGLMWATKYEVVALKD-----TLTSRPAQG 1896

QY 373 -----CVTDEC-----FFERLESNNY----- 389
DB 1897 VVTILENVSPPRARVTDATETTTISWRKTETITGFQVDAIPANGQTPVQRTISPDVR 1956

QY 390 -----NTYRSRYT-----SWYVA 403
DB 1957 SYTITGLQPGTDYKIHLYTLNDNARSPVVIDASTAIDAPNRLFLTTTPNSLLVSWQAP 2016

QY 404 LKRTQY-----KLGSK-----TG--PGOKAILFL-----PW- 428
DB 2017 RARITGYIIVKEKPGSPREVPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLI 2076

QY 429 -SAASDELQPLVTLPHPNLHGPEILDVPST 457
DB 2077 GRKTDLPQLVTLPHPNLHGPEILDVPST 2106

RESULT 4
FIND MOUSE
ID FINE MOUSE STANDARD; PRT; 2477 AA.
AC P11276; Q61567; Q61568; Q61569; Q64233; Q80U14;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fibronectin precursor (FN).
GN FN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-920 FROM N.A.
RC STRAIN=FVB/N-3; TISSUE=Breast tumor;
RX MEDLINE=23288257; PubMed=12477932;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE OF 1-28 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94131313; PubMed=82999972;
RA Polly P., Nicholson R.C.;
RT "Sequence of the mouse fibronectin-encoding gene promoter region."
RL Gene 137:353-354 (1993).
```


RN [3] SEQUENCE OF 562-834 FROM N.A.
RC STRAIN=NMRI;
RX MEDLINE=95403556; PubMed=7673336;
RA Talts J.F., Weller A., Timpl R., Ekblom M., Ekblom P.;
RT "Regulation of mesenchymal extracellular matrix protein synthesis by
transforming growth factor-beta and glucocorticoids in tumor
stroma.";
RL J. Cell Sci. 108:2153-2162(1995).
[4]
RN SEQUENCE OF 899-2376 FROM N.A.
RA Gorski G., Aros M., Norton P.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE OF 2375-2477 FROM N.A.
RX MEDLINE=88124987; PubMed=3124113;
RA Blatti S.P., Foster D.N., Ranganathan G., Moses H.L., Getz M.J.;
RT "Induction of fibronectin gene transcription and mRNA is a primary
response to growth-factor stimulation of AKR-2B cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:1119-1123(1988).
[6]
RN SEQUENCE OF 2375-2477 FROM N.A.
RT TISSUE=Kidney;
RX MEDLINE=93011702; PubMed=1327855;
RA Khandjian E.W., Salomon C., Leonard N., Tremblay S., Turler H.;
RT "Fibronectin gene expression in proliferating, quiescent, and SV40-
infected mouse kidney cells.";
RL Exp. Cell Res. 202:464-470(1992).
[7]
RN STRUCTURE BY NMR OF 1447-1630.
RX MEDLINE=96202578; PubMed=9533887;
RA Copie V., Tomita Y., Akiyama S.K., Aota S., Yamada K.M., Venable R.M.,
RT Pastor R.W., Krueger S., Torchia D.A.;
RT "Solution structure and dynamics of linked cell attachment modules of
mouse fibronectin containing the RGD and synergy regions: comparison
with the human fibronectin crystal structure.";
RL J. Mol. Biol. 277:663-682(1998).
[8]
RN DOWN-REGULATION BY GLUCOCORTICOIDS.
RX MEDLINE=21600963; PubMed=11737251;
RA Gu Y.-C., Talts J.F., Gullberg D., Timpl R., Ekblom M.;
RT "Glucocorticoids down-regulate the extracellular matrix proteins
fibronectin, fibulin-1 and fibulin-2 in bone marrow stroma.";
RL Eur. J. Haematol. 67:176-184(2001).
CC -I- FUNCTION: Fibronectins bind cell surfaces and various compounds
including collagen, fibrin, heparin, DNA, and actin. Fibronectins
are involved in cell adhesion, cell motility, opsonization, wound
healing, and maintenance of cell shape.
CC -I- SUBUNIT: Mostly heterodimers or multimers of alternatively spliced
variants, connected by 2 disulfide bonds near the carboxyl ends;
to a lesser extend homodimers. Interacts with FBLN1 (By
similarity).
CC -I- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -I- ALTERNATIVE PRODUCTS:
Comment=A number of isoforms are produced. Each of the "extra
domain" and the connecting strand 3 are present in some forms of
fibronectin and absent in others;
Name=1;
CC IsoId=P11276-1; Sequence=Displayed;
CC -I- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted
by hepatocytes. Cellular FN (dimeric or cross-linked multimeric
forms), made by fibroblasts, epithelial and other cell types, is
deposited as fibrils in the extracellular matrix.
CC -I- INDUCTION: Glucocorticoids suppressed mRNA expression and protein
synthesis.
CC -I- PTM: Sulfated (By similarity).
CC -I- SIMILARITY: Contains 12 fibronectin type I domains.
CC -I- SIMILARITY: Contains 2 fibronectin type II domains.
CC -I- SIMILARITY: Contains 17 fibronectin type III domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; BC051082; AAHS1082.1; -.
DR EMBL; Z22729; CAAB0422.1; -.
DR EMBL; X82402; CAA57796.1; -.
DR EMBL; X93167; CAA63654.1; -.
DR EMBL; M18194; AAA37636.1; -.
DR EMBL; S45680; AAB23491.1; -.
DR PIR; A49173; A49173.
DR PIR; I48349; I48349.
DR PDB; 1MFN; 29-APR-98.
DR PDB; 2MFN; 29-APR-98.
DR MGD; MGI:95566; Fnl.
DR GO; GO:0007155; Pcell adhesion; IDA.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000083; Fibrctnl.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR000562; FN_Type II.
DR InterPro; IPR003962; FnIII_subd.
DR PRINTS; PR00014; MCR_alpha_beta_C.
DR SMART; SM00058; FN1; 4.
DR SMART; SM00060; FN3; 12.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS00023; FIBRONECTIN_2; 2.
DR PROSITE; PS01253; FIBRONECTIN_1; 12.
KW Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;
KW Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal;
KW 3D-structure.
FT SIGNAL 1 32 BY SIMILARITY.
FT CHAIN 33 2477 FIBRONECTIN.
FT DOMAIN 53 273 FIBRIN- AND HEPARIN-BINDING 1.
FT DOMAIN 308 608 COLLAGEN-BINDING.
FT DNA BIND 906 1171 CELL-ATTACHMENT.
FT DOMAIN 1357 1630 HEPARIN-BINDING 2.
FT DOMAIN 1811 2081 FIBRIN-BINDING 2.
FT DOMAIN 2296 2427 FIBRONECTIN TYPE-I 1.
FT DOMAIN 51 96 FIBRONECTIN TYPE-I 2.
FT DOMAIN 96 140 FIBRONECTIN TYPE-I 3.
FT DOMAIN 140 185 FIBRONECTIN TYPE-I 4.
FT DOMAIN 185 230 FIBRONECTIN TYPE-I 5.
FT DOMAIN 230 272 FIBRONECTIN TYPE-I 6.
FT DOMAIN 306 343 FIBRONECTIN TYPE-I 7.
FT DOMAIN 345 404 FIBRONECTIN TYPE-I 8.
FT DOMAIN 405 469 FIBRONECTIN TYPE-I 9.
FT DOMAIN 468 516 FIBRONECTIN TYPE-I 10.
FT DOMAIN 516 559 FIBRONECTIN TYPE-I 11.
FT DOMAIN 559 602 FIBRONECTIN TYPE-I 12.
FT DOMAIN 609 706 FIBRONECTIN TYPE-I 13.
FT DOMAIN 707 808 FIBRONECTIN TYPE-I 14.
FT DOMAIN 809 903 FIBRONECTIN TYPE-I 15.
FT DOMAIN 904 994 FIBRONECTIN TYPE-I 16.
FT DOMAIN 995 1084 FIBRONECTIN TYPE-I 17.
FT DOMAIN 1085 1172 FIBRONECTIN TYPE-I 18.
FT DOMAIN 1173 1264 FIBRONECTIN TYPE-I 19.
FT DOMAIN 1265 1355 FIBRONECTIN TYPE-I 20.
FT DOMAIN 1356 1446 FIBRONECTIN TYPE-I 21.
FT DOMAIN 1447 1536 FIBRONECTIN TYPE-I 22.
FT DOMAIN 1537 1630 FIBRONECTIN TYPE-I 23.
FT DOMAIN 1631 1720 FIBRONECTIN TYPE-I 24.
FT DOMAIN 1721 1810 FIBRONECTIN TYPE-I 25.
FT DOMAIN 1811 1901 FIBRONECTIN TYPE-I 26.
FT DOMAIN 1903 1991 FIBRONECTIN TYPE-I 27.
FT DOMAIN 1992 2081 FIBRONECTIN TYPE-I 28.
FT DOMAIN 2082 2201 CONNECTING STRAND 3 (CS-3) (V REGION).
FT DOMAIN 2202 2283 FIBRONECTIN TYPE-I 29.
FT DOMAIN 2284 2338 FIBRONECTIN TYPE-I 30.
FT DOMAIN 2339 2381 FIBRONECTIN TYPE-I 31.
FT DOMAIN 2382 2477 FIBRONECTIN TYPE-I 32.

FT	DOMAIN	2383	2426	FIBROECTIN TYPE-I 12.
FT	SITE	1614	1616	CELL ATTACHMENT SITE.
FT	SITE	2181	2183	CELL ATTACHMENT SITE.
FT	DISULFID	53	79	BY SIMILARITY.
FT	DISULFID	77	88	BY SIMILARITY.
FT	DISULFID	98	126	BY SIMILARITY.
FT	DISULFID	124	136	BY SIMILARITY.
FT	DISULFID	142	170	BY SIMILARITY.
FT	DISULFID	168	180	BY SIMILARITY.
FT	DISULFID	187	216	BY SIMILARITY.
FT	DISULFID	214	226	BY SIMILARITY.
FT	DISULFID	232	261	BY SIMILARITY.
Query Match				
Best Local Similarity 55.9%; Score 1340; DB 1; Length 2477;				
Matches 316; Conservative 46; Mismatches 85; Indels 310; Gaps 16;				
Qy	1	PTDLRFNIGDPTWRVTWAPPSPSIDLNFVLYSPVQNEEDVABLSPSDNAVVLNLL	60	
Db	1360	PTDLRFNIGDPTWRVTWAPPSPSELTNLNVYSPVQNEEDVABLSPSDNAVVLNLL	1419	
Qy	61	PGTEYVVSVSVEQHESTPLRGQKTGLDQSPGIGDFSDITANSFTVHWIAPRATITGYR	120	
Db	1420	PGTEYLVSVSVYEQHSIPLRGQKTGLDQSPGIGDFSDITANSFTVHWIAPRATITGYI	1479	
Qy	121	IRHHPHFSGRPREDRVPHRSNITLNLTPGTETVYVSVIYALNGREESPLLIGQOSTVSD	180	
Db	1480	IRHHAHSVGRPRQDRVPPSRNSITLNLNPGTEYVSVIYALNGREESPLLIGQATVSD	1539	
Qy	181	VPRDLVVAATPTSLLSISWAPAVTVRYRITYGETGNSPVQBFVPGSKSTATISGLK	240	
Db	1540	IPRDLVIAATPTSLLSISWEPVAVSVRYRITYGETGNSPVQBFVPGSKSTATINIK	1599	
Qy	241	PGVDYTTVAVTGRGDSPASSKDISNYREIDKPS-----M	278	
Db	1600	PGADYTTLVAVTGRGDSPASSKPSVINYKTEIDKPSQMOMQVTDVODNSISVRMLPSTSPV	1659	
Qy	279	AAGSITLTP-----	287	
Db	1660	TGYRVTTTPKNGLGPSKTKTASPDQETMTIEGLQPTVEYVSVYAQNNGESQPLVQTAV	1719	
Qy	288	-----ALPEDGGSGAPPQPHFKD-----	305	
Db	1720	TNIDRPKGLAFTDVDVDSIKIAMESPQGVSRVRYTSSPDGIRELFPAPDGEDDTAEL	1779	
Qy	306	-----PKELYCKN-----	313	
Db	1780	QGLRPGSEYTVSVVALHDDMESQPLIGIQSTAIPTNLKLQSVTPTSTFAQWIAPSVQL	1839	
Qy	314	GGFPLRTHPDGRVGDVREKSDPHIKLQQAEEGWGSIKGVCANRY---LAKMEDGRLL	369	
Db	1840	TGYRVRYNP-----KEKTPMKEINLSPDSSVIVGLMVAKYEVSVYALKD---TL	1889	
Qy	370	ASK-----CVTDEC-----FFFERLESNNY-----	389	
Db	1890	TSRPAQGVITTLNVSPPRARVTDATETTTISWRKTETITGFQVDAIPANGQTPVQR	1949	
Qy	390	-----NTRYRSRYT-----	398	
Db	1950	SISPDVRSYTTIGLQPTGDKYKIHLYTLNDNARSPVIIDASTAIDAPSNLRFLLTTPNSL	2009	
Qy	399	--SWYVALKRTGQY-----KLGSK-----TG--PGOKAILFL-----	426	
Db	2010	LVSHQAAPRATITGYIIKYEPGSPPREVPRPRPGVTEATITGLEPGETIYIVIALKNN	2069	
Qy	427	----PM--SAASDELQPLVTLPHNPHLHGPEILDVPST	457	
Db	2070	QKSEPLIGRKKTDELQPLVTLPHNPHLHGPEILDVPST	2106	
RESULT 5				
FINC CHICK				
ID	FINC CHICK	STANDARD;	PRT;	1256 AA.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

	EMBL; V00432; CAA23714.1; -	
DR	EMBL; U21327; AAA73566.1; -	
DR	EMBL; X06533; CAA29781.1; -	
DR	EMBL; M26186; ABA48772.1; ALT_SEQ.	
DR	EMBL; U20386; BAB01062.1; -	
DR	PIR; A28512; A28512.	
DR	PIR; A29355; A29355.	
DR	PIR; S71465; S71465.	
DR	HSSP; P02751; IFNH.	
DR	InterPro; IPRO000083; Fibronctnl.	
DR	InterPro; IPRO08957; FN_III-like.	
DR	InterPro; IPRO03961; FN_III.	
DR	InterPro; IPRO00562; FN_Type_II.	
DR	InterPro; IPRO03962; FNIii subd.	
DR	PRINTS; PR00014; FNTYPEI.II.	
DR	SMART; SM00058; FNI; 3.	
DR	SMART; SM00060; FN3; 9.	
DR	PROSITE; PS00023; FIBONECTIN 2; PARTIAL.	
DR	PROSITE; PS01253; FIBONECTIN 1; 2.	
KW	Glycoprotein; Plasma; Heparin-binding; Acute phase; Cell adhesion;	
KW	Repeat; Sulfation; Alternative splicing.	
FT	NON_TER	1
FT	NON_CONS	50
FT	DOMAIN	236 509
FT	DOMAIN	690 961
FT	DOMAIN	1153 1226
FT	DOMAIN	327 415
FT	DOMAIN	416 509
FT	DOMAIN	510 599
FT	DOMAIN	600 689
FT	DOMAIN	690 781
FT	DOMAIN	782 871
FT	DOMAIN	872 961
FT	DOMAIN	962 1082
FT	DOMAIN	1083 1152
FT	DOMAIN	1174 1218
FT	DOMAIN	1219 >1256
FT	SITE	493 495
FT	DISULFID	1176 1205
FT	DISULFID	1203 1215
FT	DISULFID	1221 1248
FT	CARBOHYD	122 122
FT	CARBOHYD	1078 1078
FT	CARBOHYD	1034 1034
FT	CARBOHYD	1035 1035
FT	CARBOHYD	516 516
FT	CONFLICT	569 572
FT	NON_TER	1256 1256
FO	SEQUENCE	1256 AA; 137435 MW; 345AA4CA0B4D71D9B CRC64;

[illegible]

```
FT DOMAIN 202 292 FIBRONECTIN TYPE-III.
FT DOMAIN 383 383 FIBRONECTIN TYPE-III.
FT DOMAIN 394 477 FIBRONECTIN TYPE-III.
FT DOMAIN 478 567 FIBRONECTIN TYPE-III.
FT DOMAIN 568 657 FIBRONECTIN TYPE-III.
FT DOMAIN 658 749 FIBRONECTIN TYPE-III.
FT DOMAIN 750 838 FIBRONECTIN TYPE-III.
FT DOMAIN 839 929 FIBRONECTIN TYPE-III.
FT DOMAIN 930 1064 CONNECTING STRAND 3 (CS-3) (V REGION).
FT DOMAIN 1065 1143 FIBRONECTIN TYPE-III.
FT DOMAIN 1155 1199 FIBRONECTIN TYPE-I.
FT DOMAIN 1200 1243 FIBRONECTIN TYPE-I.
FT DOMAIN 1245 1287 FIBRONECTIN TYPE-I.
FT DISULFID 1157 1186 BY SIMILARITY.
FT DISULFID 1184 1196 BY SIMILARITY.
FT DISULFID 1202 1229 BY SIMILARITY.
FT DISULFID 1227 1240 BY SIMILARITY.
FT DISULFID 1247 1270 BY SIMILARITY.
FT DISULFID 1268 1284 BY SIMILARITY.
FT DISULFID 1320 1320 INTERCHAIN (WITH C-1324) (BY SIMILARITY).
FT DISULFID 1324 1324 INTERCHAIN (WITH C-1320) (BY SIMILARITY).
FT SITE 461 463 CELL ATTACHMENT SITE.
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1328 AA; 145037 MW; E31BF7968AID1E74 CRC64;

Query Match 43.8%; Score 1051.5; DB 1; Length 1328;
Best Local Similarity 71.1%; Pred. No. 1.9e-66;
Matches 197; Conservative 41; Mismatches 38; Indels 1; Gaps 1;

QY 2 TDLRFTNIGDTRVTPWAPPSSIDLTLNVLRYSPVKNEDVAELSPSDNVAVLTLNLP 61
Db 207 TDLRFTNVGPDMLVTSAPPSSVLSFLRVYVPSKNEEDAAELTSPSDNVAVLTLNLP 266
QY 62 GTEYVVSVSIVYQHSTPLRGKQGLDPTGIDFSDITANSFTVHWIAPRATITGYRI 121
Db 267 GTEYIVSVFAVYERESTPLTGVTQRTGLDPTGIDFSDITSSFTVYVWAPRATITGYKI 326
QY 122 RHPEFSGRPREDR-VPHSRNSITLTLNTPGTYVVSIVALNGREESPLLICQOSTVSD 180
Db 327 QYHPETGGAGQKERCVCVPSRNSLTTLNTPGTYVVSIVAFVNGRESVPLVQQATVSD 386
QY 181 VPRDLEVVAAPTSLISWDAPAVTVYRITTYGETGNSPVQEFVTPGSKSTATISGLK 240
Db 387 TPTNLEVTSTPTSMISWDAPVGVYRITTYTGTGTPVQEFVTPGDRSDAPIRGLK 446
QY 241 PGVDYITVYVAVTGRGDSPASSKPIISINXRTEDKPS 277
Db 447 PGAEXIITVYVAVTGRGDSPASSKPIVTHKTVVDKPT 483
```

RESULT 7

```
FINC_XENLA STANDARD; PRT; 2481 AA.
ID -FINC_XENLA STANDARD; PRT; 2481 AA.
AC Q91740;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Fibrinectin precursor.
GN FN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92111942; PubMed=1730390;
RA Desimone D.W., Norton P.A., Hynes R.O.;
RT "Identification and characterization of alternatively spliced
RL fibrinectin mRNAs expressed in early Xenopus embryos.";
RL Dev. Biol. 149:357-369(1992).
CC -!- FUNCTION: Fibrinectins bind cell surfaces and various compounds
CC including collagen, fibrin, heparin, DNA, and actin. Fibrinectins
```

```
CC are involved in cell adhesion, cell motility, opsonization, wound
CC healing, and maintenance of cell shape (By similarity).
CC -!- SUBUNIT: DIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED VARIANTS,
CC CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS (BY
CC SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. Each of the "extra
CC domain" and the connecting strand 3 are present in some forms of
CC fibrinectin and absent in others;
CC Name=1;
CC IsoId=Q91740-1; Sequence=Displayed; EMBRYO, CELLULAR FORMS OF
CC -!- TISSUE SPECIFICITY: IN EARLY XENOPUS EMBRYO, CELLULAR FORMS OF
CC FIBRONECTIN PREDOMINATE WHICH INCLUDE BOTH EXTRA DOMAINS. IN
CC FIBRONECTIN OF EMBRYONIC AND ADULT LIVER THE CONNECTING STRAND 3
CC CAN BE EITHER COMPLETELY EXCLUDED OR INCLUDED.
CC -!- SIMILARITY: Contains 12 fibrinectin type I domains.
CC -!- SIMILARITY: Contains 2 fibrinectin type II domains.
CC -!- SIMILARITY: Contains 17 fibrinectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M77820; AAA49707.1; -.
CC HSSP; P02751; 2FN2.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR000083; Fibrinctn1.
CC InterPro; IPR008957; FN III-like.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR000562; FN_Type_II.
CC Pfam; PF00039; fn1; 12.
CC Pfam; PF00040; fn2; 2.
CC Pfam; PF00041; fn3; 17.
CC PRINTS; PR00013; FNTYPEII.
CC ProDom; PD000995; FN_Type_II; 2.
CC SMART; SM00058; FN1; 12.
CC SMART; SM00059; FN2; 2.
CC SMART; SM00060; FN3; 17.
CC PROSITE; PS00022; EGF 1; 2.
CC PROSITE; PS01253; FIBRONECTIN 1; 11.
CC Glycoprotein; Plasma; Heparin-binding; Acute phase; Cell adhesion;
CC Repeat; Alternative splicing; Signal
CC SIGNAL 1 31
CC FIBRONECTIN.
CC CHAIN 32 2481
CC FT DOMAIN 55 275 FIBRIN- AND HEPARIN-BINDING 1.
CC FT DOMAIN 309 609 COLLAGEN-BINDING.
CC FT DNA_BIND 907 1172 BY SIMILARITY.
CC FT DOMAIN 1358 1631 CELL-ATTACHMENT.
CC FT DOMAIN 1812 2082 HEPARIN-BINDING 2.
CC FT DOMAIN 2301 2432 FIBRIN-BINDING 2.
CC FT DOMAIN 53 93 FIBRONECTIN TYPE-I 1.
CC FT DOMAIN 98 141 FIBRONECTIN TYPE-I 2.
CC FT DOMAIN 142 185 FIBRONECTIN TYPE-I 3.
CC FT DOMAIN 187 231 FIBRONECTIN TYPE-I 4.
CC FT DOMAIN 232 276 FIBRONECTIN TYPE-I 5.
CC FT DOMAIN 307 346 FIBRONECTIN TYPE-I 6.
CC FT DOMAIN 346 405 FIBRONECTIN TYPE-II 1.
CC FT DOMAIN 406 470 FIBRONECTIN TYPE-II 2.
CC FT DOMAIN 469 512 FIBRONECTIN TYPE-I 7.
CC FT DOMAIN 517 559 FIBRONECTIN TYPE-I 8.
CC FT DOMAIN 560 603 FIBRONECTIN TYPE-I 9.
CC FT DOMAIN 610 707 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 708 809 FIBRONECTIN TYPE-III 2.
CC FT DOMAIN 810 904 FIBRONECTIN TYPE-III 3.
CC FT DOMAIN 905 995 FIBRONECTIN TYPE-III 4.
CC FT DOMAIN 996 1085 FIBRONECTIN TYPE-III 5.
CC FT DOMAIN 1086 1173 FIBRONECTIN TYPE-III 6.
CC FT DOMAIN 1174 1265 FIBRONECTIN TYPE-III 7.
```

```
FT DOMAIN 1266 1356 FIBRONECTIN TYPE-III 8 (EXTRA DOMAIN).
FT DOMAIN 1357 1447 FIBRONECTIN TYPE-III 9.
FT DOMAIN 1448 1537 FIBRONECTIN TYPE-III 10.
FT DOMAIN 1538 1631 FIBRONECTIN TYPE-III 11.
FT DOMAIN 1632 1721 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1722 1811 FIBRONECTIN TYPE-III 13 (EXTRA DOMAIN).
FT DOMAIN 1812 1903 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1904 1992 FIBRONECTIN TYPE-III 15.
FT DOMAIN 1993 2082 FIBRONECTIN TYPE-III 16.
FT DOMAIN 2083 2205 CONNECTING STRAND 3 (CS-3) (V REGION).
FT DOMAIN 2206 2287 FIBRONECTIN TYPE-III 17.
FT DOMAIN 2288 2343 FIBRONECTIN TYPE-I 10.
FT DOMAIN 2344 2386 FIBRONECTIN TYPE-I 11.
FT DOMAIN 2388 2431 FIBRONECTIN TYPE-I 12.
FT SITE 1615 1617 CELL ATTACHMENT SITE.
FT DISULFID 55 81 BY SIMILARITY.
FT DISULFID 79 90 BY SIMILARITY.
FT DISULFID 100 128 BY SIMILARITY.
FT DISULFID 126 138 BY SIMILARITY.
FT DISULFID 144 172 BY SIMILARITY.
FT DISULFID 170 182 BY SIMILARITY.
FT DISULFID 189 218 BY SIMILARITY.
FT DISULFID 216 228 BY SIMILARITY.
FT DISULFID 234 263 BY SIMILARITY.
FT DISULFID 261 273 BY SIMILARITY.
FT DISULFID 309 336 BY SIMILARITY.
FT DISULFID 334 343 BY SIMILARITY.
FT DISULFID 361 387 BY SIMILARITY.
FT DISULFID 375 402 BY SIMILARITY.
FT DISULFID 421 447 BY SIMILARITY.
FT DISULFID 471 499 BY SIMILARITY.
FT DISULFID 497 509 BY SIMILARITY.
FT DISULFID 519 546 BY SIMILARITY.
FT DISULFID 544 556 BY SIMILARITY.
FT DISULFID 562 590 BY SIMILARITY.
FT DISULFID 588 600 BY SIMILARITY.
FT DISULFID 2301 2330 BY SIMILARITY.
FT DISULFID 2328 2340 BY SIMILARITY.
FT DISULFID 2346 2373 BY SIMILARITY.
FT DISULFID 2371 2383 BY SIMILARITY.
FT DISULFID 2390 2414 BY SIMILARITY.
FT DISULFID 2412 2428 BY SIMILARITY.
FT DISULFID 2459 2459 INTERCHAIN (WITH C-2463) (BY SIMILARITY).
FT DISULFID 2463 2463 INTERCHAIN (WITH C-2459) (BY SIMILARITY).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 543 543 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 877 877 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1244 1244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1291 1291 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2202 2202 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 2481 AA; 272678 MW; 7E47DF4F6CE72C93 CRC64;
Query Match 42.6%; Score 1020.5; DB 1; Length 2481;
Best Local Similarity 68.7%; Pred. No. 6.9e-64;
Matches 195; Conservative 44; Mismatches 44; Indels 1; Gaps 1;
QY 1 PTDLRFNIGDPTMTVWAPPSPDLTNFLVRYSPVKNEDVAELSPSDNAVLTNLL 60
DB 1361 PTLNRFNIGDPTMTVWAPPSPDLTNFLVRYSPVKNEDVAELSPSDNAVLTNLL 1420
QY 61 PGTEYVSVSVSEHQESTPLRGQKGTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 120
DB 1421 PFTYLVSVSVSEHQESTPLRGQKGTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 1480
QY 121 IRHHPHFSGRPREDRVPHSRNSTLTNLTPTGTEYVSVSVSEHQESTPLRGQKGTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 180
DB 1481 IRYQLSGAGRPKEERVPSPRSNTLTNLTPTGTEYVSVSVSEHQESTPLRGQKGTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 1540
QY 181 VPRDLVVAATPTSLISWADAPATVYRYITVGTGCGSPVQFVTPGSKSTATISGLK 240
DB 1541 VPTDLEVTSSPNTLTISWADAPATVYRYITVGTGCGSPVQFVTPGSKSTATISGLK 1600
```

```
QY 241 PGVDYITVYAVTGRGDSPASSKPSISINVTETDKP-SMAAGSI 283
DB 1601 PGVSYTITVYAVTGRGDSPASSKPSLTIIHKTDVDQPIDMAVTDI 1644
RESULT 8
FGF2_HUMAN STANDARD; PRT; 155 AA.
ID FGF2_HUMAN AC P09038;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast growth factor) (BFGF) (Prostatropin).
GN FGF2 OR FGF8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87053817; PubMed=3780670;
RA Abraham J.A., Whang J.L., Tumolo A., Mergia A., Friedman J.,
RA Gospodarowicz D., Fiddes J.C.;
RT "Human basic fibroblast growth factor: nucleotide sequence and genomic organization."
RL EMBO J. 5:2523-2528(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87217066; PubMed=3472745;
RA Abraham J.A., Whang J.L., Tumolo A., Mergia A., Fiddes J.C.;
RT "Human basic fibroblast growth factor: nucleotide sequence, genomic organization, and expression in mammalian cells."
RL Cold Spring Harb. Symp. Quant. Biol. 51:657-668(1986).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87213238; PubMed=3579930;
RA Sommer A., Brewer M.T., Thompson R.C., Moscatelli D., Presta M.,
RA Rifkin D.B.;
RT "A form of human basic fibroblast growth factor with an extended amino terminus."
RL Biochem. Biophys. Res. Commun. 144:543-550(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=87162468; PubMed=2435575;
RA Kurokawa T., Sasada R., Iwane M., Igarashi K.;
RT "Cloning and expression of cDNA encoding human basic fibroblast growth factor."
RL FEBS Lett. 213:189-194(1987).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=89184522; PubMed=2539817;
RA Prats H., Kaghad M., Prats A.C., Klagesbrun M., Lelias J.M.,
RA Liauzun P., Chalou P., Tauber J.P., Amalric F., Smith J.A.,
RA Caput D.;
RT "High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons."
RL Proc. Natl. Acad. Sci. U.S.A. 86:1836-1840(1989).
RN [6]
RP SEQUENCE OF 10-35.
RX MEDLINE=86275260; PubMed=3732516;
RA Gautschi P., Frater-Schroeder M., Boehlen P.;
RT "Partial molecular characterization of endothelial cell mitogens from human brain: acidic and basic fibroblast growth factors."
RL FEBS Lett. 204:203-207(1986).
RN [7]
RP SEQUENCE OF 10-39.
RX MEDLINE=86186784; PubMed=3964259;
RA Gimenez-Gallego G., Conn G., Hatcher V.B., Thomas K.A.;
RT "Human brain-derived acidic and basic fibroblast growth factors: amino terminal sequences and specific mitogenic activities."
RL Biochem. Biophys. Res. Commun. 135:541-548(1986).
RN [8]
```

SEQUENCE OF 2-22.
RX MEDLINE=87156686; PubMed=2435284;
RA Story M.T., Esch F., Shimasaki S., Sasse J., Jacobs S.C., Lawson R.K.;
RT "Amino-terminal sequence of a large form of basic fibroblast growth
factor isolated from human benign prostatic hyperplastic tissue."
RL Biochem. Biophys. Res. Commun. 142:702-709(1987).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=92121151; PubMed=1769963;
RA Ago H., Kitagawa Y., Fujishima A., Matsuura Y., Katsube Y.;
RT "Crystal structure of basic fibroblast growth factor at 1.6-A
resolution."
RL J. Biochem. 110:360-363(1991).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=91195367; PubMed=1707542;
RA Eriksson A.E., Cousins L.S., Weaver L.H., Matthews B.W.;
RT "Three-dimensional structure of human basic fibroblast growth
factor."
RL Proc. Natl. Acad. Sci. U.S.A. 88:3441-3445(1991).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=91195368; PubMed=1849658;
RA Zhang J., Cousins L.S., Barr P.J., Sprang S.R.;
RT "Three-dimensional structure of human basic fibroblast growth factor,
a structural homolog of interleukin 1 beta."
RL Proc. Natl. Acad. Sci. U.S.A. 88:3446-3451(1991).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=91095983; PubMed=1702556;
RA Zhu X., Komiya H., Chirino A., Faham S., Fox G.M., Arakawa T.,
Heu B.T., Rees D.C.;
RT "Three-dimensional structures of acidic and basic fibroblast growth
factors."
RL Science 251:90-93(1991).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=9400464; PubMed=7691311;
RA Eriksson A.E., Cousins L.S., Matthews B.W.;
RT "Refinement of the structure of human basic fibroblast growth factor
at 1.6-A resolution and analysis of presumed heparin binding sites by
selenate substitution."
RL Protein Sci. 2:1274-1284(1993).
RN [14]
RP STRUCTURE BY NMR.
RX MEDLINE=97040521; PubMed=8895834;
RA Moy F.J., Seddon A.P., Boehlen P., Powers R.;
RT "High-resolution solution structure of basic fibroblast growth factor
determined by multidimensional heteronuclear magnetic resonance
spectroscopy."
RL Biochemistry 35:13552-13561(1996).
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
in vivo and are potent mitogens for a variety of cell types in
vitro. There are differences in the tissue distribution and
concentration of these 2 growth factors.
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
aFGF.
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17599; AAA52534.1; ALT_INIT.
DR EMBL; X04431; CAA28027.1; -.
DR EMBL; X04432; CAA28028.1; -.
DR EMBL; X04433; CAA28029.1; -.
DR EMBL; M27968; AAA52448.1; -.

DR EMBL; J04513; AAA52533.1; ALT_INIT.
DR PIR; A32398; A32398.
DR PDB; 2FGF; 15-APR-92.
DR PDB; 4FGF; 15-JUL-93.
DR PDB; 1FGA; 15-JUL-93.
DR PDB; 1BFB; 03-APR-96.
DR PDB; 1BFC; 03-APR-96.
DR PDB; 1BFF; 16-JUN-97.
DR PDB; 1BFG; 31-JAN-94.
DR PDB; 2BFH; 30-APR-94.
DR PDB; 1BLA; 08-NOV-96.
DR PDB; 1BLD; 08-NOV-96.
DR PDB; 1BAS; 31-OCT-93.
DR PDB; 1CVS; 28-JAN-00.
DR PDB; 1EV2; 31-MAY-00.
DR PDB; 1FQ9; 27-SEP-00.
DR PDB; 11I4; 20-JUN-01.
DR PDB; 11IL; 20-JUN-01.
DR Genew; HGNC:3676; FGF2.
DR MIM; 134920; -.
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0000187; P:activation of MAPK; TAS.
DR GO; GO:0006935; P:chemotaxis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR GO; GO:0008284; P:neurogenesis; TAS.
DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
DR GO; GO:0007265; P:RAS protein signal transduction; TAS.
DR InterPro; IPR008996; Cytok_IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR ProDom; PD000831; IL1_HBGF; 1.
DR SMART; SM00443; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding;
KW 3D-structure.
FT PROPEP 1 9
FT CHAIN 10 155 HEPARIN-BINDING GROWTH FACTOR 2.
FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 88 90 CELL ATTACHMENT SITE (POTENTIAL).
FT BINDING 27 31 HEPARIN (POTENTIAL).
FT BINDING 116 119 HEPARIN (POTENTIAL).
FT STRAND 30 34 HEPARIN (POTENTIAL).
FT TURN 35 38
FT STRAND 39 43
FT TURN 45 46
FT STRAND 49 52
FT TURN 55 56
FT TURN 58 60
FT STRAND 62 66
FT TURN 69 70
FT STRAND 71 76
FT TURN 77 80
FT STRAND 81 85
FT TURN 87 88
FT TURN 91 94
FT STRAND 99 101
FT STRAND 103 107
FT TURN 109 110
FT STRAND 113 117
Query Match 34.2%; Score 820; DB 1; Length 155;
Best Local Similarity 99.4%; Pred. No. 2.2e-51;
Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 278 MAAGSITLPAIPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHI 337
DB 1 MAAGSITLPAIPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHI 60
QY 338 KIQLOAERGVSIVKVCANRYLAMKEDGRLLASKCVTDECFFERLSSNNYTSRSKY 397
DB 61 KIQLOAERGVSIVKVCANRYLAMKEDGRLLASKCVTDECFFERLSSNNYTSRSKY 120


```
QY 398 TSWYVALKRTGQYKLGSKTGPQKAILFLPMSAAS 432
:|||||
Db 121 SSWYVALKRTGQYKLGSKTGPQKAILFLPMSAKS 155

RESULT 10
FGF2_SHEEP
ID FGF2_SHEEP STANDARD; PRT; 155 AA.
AC P20003;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF) (Prostatropin).
GN FGF2 OR FGF-2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Sutton R., Ward W.G., Raphael K.A., Cam G.R.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 9-155.
RX MEDLINE=8055577; PubMed=3678486;
RA Simpson R.J., Moritz R.L., Lloyd C.J., Fabri L.J., Nice E.C.,
RA Rubira M.R., Burgess A.W.;
RT "Primary structure of ovine pituitary basic fibroblast growth
RT factor.";
RT FEBS Lett. 224:128-132(1987).
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in
CC vitro. There are differences in the tissue distribution and
CC concentration of these 2 growth factors.
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC aFGF.
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L36136; AAA31519.1; -.
DR HSP; P09038; 1BFF.
DR InterPro; IPR008996; Cytok IL1-like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PD00262; IL1HBGF.
DR PRODom; PD000831; IL1_HBGF; 1.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT PROPEP 1 9
FT CHAIN 10 155 HEPARIN-BINDING GROWTH FACTOR 2.
FT SITE 45 48 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 87 90 CELL ATTACHMENT SITE (POTENTIAL).
FT BINDING 27 31 HEPARIN (POTENTIAL).
FT BINDING 116 119 HEPARIN (POTENTIAL).
FT BINDING 155 AA; 17280 MW; B5F2364BA610606D CRC64;
SQ

Query Match 33.6%; Score 805; DB 1; Length 155;
Best Local Similarity 97.4%; Pred. No. 2.6e-50;
Matches 151; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 278 MAAGSITTLPALPEDGSGAFPPGHFKDPKRLYCKNGGFFLRIHPDGRVGVREKSDPHI 337
:|||||
```

```
Db 1 MAAGSITTLPALPEDGSGAFPPGHFKDPKRLYCKNGGFFLRIHPDGRVGVREKSDPHI 60
338 KLQLQAEERGVVSIKGVCANRYLAMKEDGRLLASKCVTDECFFERLESNNYNTYRSKY 397
|||||
Db 61 KLQLQAEERGVVSIKGVCANRYLAMKEDGRLLASKCVTDECFFERLESNNYNTYRSKY 120
|||||

398 TSWYVALKRTGQYKLGSKTGPQKAILFLPMSAAS 432
:|||||
Db 121 SSWYVALKRTGQYKLGSKTGPQKAILFLPMSAKS 155

RESULT 11
FGF2_RAT
ID FGF2_RAT STANDARD; PRT; 154 AA.
AC P13109;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF) (Prostatropin).
GN FGF2 OR FGF-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Ovary;
RX MEDLINE=89061721; PubMed=3196337;
RA Shimasaki S., Emoto N., Koba A., Mercado M., Shibata F.,
RA Cooksey K., Baird A., Ling N.;
RT "Complementary DNA cloning and sequencing of rat ovarian basic
RT fibroblast growth factor and tissue distribution study of its mRNA.";
RL Biochem. Biophys. Res. Commun. 157:256-263(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88262516; PubMed=3387229;
RA Kurokawa T., Seno M., Igarashi K.;
RT "Nucleotide sequence of rat basic fibroblast growth factor cDNA.";
RL Nucleic Acids Res. 16:5201-5201(1988).
RN [3]
RP SEQUENCE OF 1-28 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Testis;
RX MEDLINE=97200905; PubMed=9048734;
RA Pasumarthi K.B.S., Jin Y., Cattini P.A.;
RT "Cloning of the rat fibroblast growth factor-2 promoter region and
RT its response to mitogenic stimuli in glioma C6 cells.";
RL J. Neurochem. 68:898-908(1997).
RN [4]
RP SEQUENCE OF 35-154 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=92329546; PubMed=1378302;
RA El-Husseini A.E.-D., Paterson J.A., Myal Y., Shiu R.P.C.;
RT "PCR detection of the rat brain basic fibroblast growth factor (bFGF)
RT mRNA containing a unique 3' untranslated region.";
RL Biochim. Biophys. Acta 1131:314-316(1992).
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in
CC vitro. There are differences in the tissue distribution and
CC concentration of these 2 growth factors.
CC -!- SUBUNIT: Monomer.
CC -!- TISSUE SPECIFICITY: Found in all tissues examined.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC aFGF.
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
```



```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22427; AAA41210.1; -.
DR EMBL; X70285; CAA30265.1; -.
DR EMBL; U78079; AAC53225.1; -.
DR EMBL; X61697; CAA43863.1; -.
DR PIR; A31674; A31674.
DR HSSP; P09038; 1BFF.
DR InterPro; IPR008996; Cytok IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR PRINTS; PR00167; FGF; 1.
DR PRODOM; PD000831; IL1_HBGF.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
DR Growth factor; Mitogen; Angiogenesis; Heparin-binding.
KW PROPEP 1
FT CHAIN 10 154 HEPARIN-BINDING GROWTH FACTOR 2.
FT BINDING 26 30 HEPARIN (POTENTIAL).
FT BINDING 115 118 HEPARIN (POTENTIAL).
SQ SEQUENCE 154 AA; 17139 MW; 1A0F14FF423D8403 CRC64;

Query Match 33.0%; Score 792.5; DB 1; Length 154;
Best Local Similarity 96.1%; Pred. No. 1.9e-49;
Matches 149; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 278 MAAGSITTLPALPEDGGGAFPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHI 337
DB 1 MAAGSITSLPALPEDGG-GAAPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHV 59

QY 338 KLOQAEERGVSIGVCANRYLAMKEDGRLLASKVCVTDECFFERLESNNYNTYRSKY 397
DB 60 KLOQAEERGVSIGVCANRYLAMKEDGRLLASKVCVTDECFFERLESNNYNTYRSKY 119

QY 398 TSMYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
DB 120 SSMYVALKRTGYKLGSKTGPQKAILFLPMSAKS 154

RESULT 12
FGF2 MOUSE
ID_FGF2_MOUSE STANDARD; PRT; 154 AA.
AC P15655;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF) (Prostatropin).
GN FGF2 OR FGF-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90201563; PubMed=23118343;
RA Hebert J.M., Basiglio C., Goldfarb M., Haub O., Martin G.R.;
RT "Isolation of cDNAs encoding four mouse FGF family members and
RT characterization of their expression patterns during embryogenesis.";
RL Dev. Biol. 138:454-463 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, A/J, and NOD/LtJ; TISSUE=Spleen;
RA Ma R.Z., Teuscher C.;
RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in
CC vitro. There are differences in the tissue distribution and
CC concentration of these 2 growth factors.
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC aFGF.
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
```

```
CC      vitro. There are differences in the tissue distribution and
CC      concentration of these 2 growth factors.
CC      -!- SUBUNIT: Monomer.
CC      -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC      aFGF.
CC      -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; Z15154; CAA78854.1; ALT_INIT.
CC      HSSP; P09038; 1BFF.
CC      InterPro; IPR008996; Cytok_IL1_like.
CC      InterPro; IPR002348; IL1_HBGF.
CC      Pfam; PF00167; FGF; 1.
CC      PRINTS; PR00262; IL1HBGF.
CC      ProDom; PD000831; IL1_HBGF; 1.
CC      SMART; SM00442; FGF; 1.
CC      PROSITE; PS00247; HBGF_FGF; 1.
CC      Growth factor; Mitogen; Angiogenesis; Heparin-binding.
CC      KW PROPEP 1 9 BY SIMILARITY.
CC      FT CHAIN 10 156 HEPARIN-BINDING GROWTH FACTOR 2.
CC      FT BINDING 28 32 HEPARIN (POTENTIAL).
CC      FT BINDING 117 120 HEPARIN (POTENTIAL).
CC      SQ SEQUENCE 156 AA; 17303 MW; 7E655FCC49BF1209 CRC64;
CC      -----
CC      Query Match 31.5%; Score 754.5; DB 1; Length 156;
CC      Best Local Similarity 92.3%; Pred. No. 9.7e-47;
CC      Matches 144; Conservative 5; Mismatches 6; Indels 1; Gaps 1;
CC      -----
QY 278 MAAGSITTLPALPED-GSGAGFPFGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPH 336
Db      |||||
Db 1 MAAGSITTLPALSDGGGGGAGFPFGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPH 60
QY 337 IKLQAEERGVVSIKGVCANRYLWAMKEDGRLLAKCVTDECFERLESNNYTYRSRK 396
Db      |||||
Db 61 IKLQAEERGVVSIKGVCANRYLWAMKEDGRLLAKCVTDECFERLESNNYTYRSRK 120
QY 397 YTSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
Db      |||||
Db 121 YSNWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 156
CC      -----
RESULT 14
FGF2_CHICK STANDARD; PRT; 158 AA.
AC P48800;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF).
GN FGF2 OR FGF-2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]_SEQUENCE FROM N.A.
RP MEDLINE=93246053; PubMed=7683281;
RA Borja A.Z., Zeller R., Weijers C.;
RT "Expression of alternatively spliced bFGF first coding exons and
RT antisense mRNAs during chicken embryogenesis.";
RL Dev. Biol. 157:110-118(1993).
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in
CC vitro. There are differences in the tissue distribution and
```

```
CC      concentration of these 2 growth factors.
CC      -!- SUBUNIT: Monomer.
CC      -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC      aFGF.
CC      -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M95707; AAA48617.1; -.
CC      HSSP; P09038; 1BFF.
CC      InterPro; IPR008996; Cytok_IL1_like.
CC      InterPro; IPR002348; IL1_HBGF.
CC      Pfam; PF00167; FGF; 1.
CC      PRINTS; PR00262; IL1HBGF.
CC      ProDom; PD000831; IL1_HBGF; 1.
CC      SMART; SM00442; FGF; 1.
CC      PROSITE; PS00247; HBGF_FGF; 1.
CC      Growth factor; Mitogen; Angiogenesis; Heparin-binding.
CC      KW PROPEP 1 12 BY SIMILARITY.
CC      FT CHAIN 13 158 HEPARIN-BINDING GROWTH FACTOR 2.
CC      FT BINDING 30 34 HEPARIN (POTENTIAL).
CC      FT BINDING 119 122 HEPARIN (POTENTIAL).
CC      SQ SEQUENCE 158 AA; 17374 MW; 7B69B684C17F1816 CRC64;
CC      -----
CC      Query Match 31.4%; Score 753; DB 1; Length 158;
CC      Best Local Similarity 91.6%; Pred. No. 1.3e-46;
CC      Matches 141; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
CC      -----
QY 279 AAGSITTLPALPEDGGGAGFPFGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIK 338
Db      |||||
Db 5 AAGSITTLPALPDGGGGAGFPFGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIK 64
QY 339 IQLQAEERGVVSIKGVCANRYLWAMKEDGRLLAKCVTDECFERLESNNYTYRSKYT 398
Db      |||||
Db 65 IQLQAEERGVVSIKGVCANRYLWAMKEDGRLLAKCVTDECFERLESNNYTYRSKYS 124
QY 399 SWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
Db      |||||
Db 125 DWYVALKRTGYKGPCKTGPQKAILFLPMSAKS 158
CC      -----
RESULT 15
FGF2_RABIT STANDARD; PRT; 137 AA.
AC P48759;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 (HBGF-2) (Basic fibroblast growth
DE factor) (BFGF) (Prostatropin) (Fragment).
GN FGF2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=New Zealand white; TISSUE=Smooth muscle;
RX MEDLINE=93343209; PubMed=8342599;
RA Winkles J.A., Friesel R., Alberts G.F., Janat M.F., Liao G.;
RT "Elevated expression of basic fibroblast growth factor in an
RT immortalized rabbit smooth muscle cell line.";
RL Am. J. Pathol. 143:518-527(1993).
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in
CC vitro. There are differences in the tissue distribution and
CC concentration of these 2 growth factors.
```

```
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC aFGF.
CC
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L12034; AAA31248.1; -.
DR PIR; I46711; I46711.
DR HSSP; P09038; 1BFF.
DR InterPro; IPR008996; Cytok IL1 like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR PRODOM; PD000831; IL1_HBGF; 1.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT BINDING 18 22
FT BINDING 107 110 HEPARIN (POTENTIAL).
FT NON TER 137 137
SQ SEQUENCE 137 AA; 15418 MW; 0D9EE457B88E8C51 CRC64;

Query Match 30.8%; Score 738; DB 1; Length 137;
Best Local Similarity 99.3%; Pred. No. 1.2e-45;
Matches 136; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 287 PALPEDGGSGAFPPGHPKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEER 346
Db 1 PALPEDGGSGAFPPGHPKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEER 60
Qy 347 GWSIKGVCANRYLAMKEDGRLASKCVTDECFFERLESNNYNTYRSKYTSWYVALKR 406
Db 61 GWSIKGVCANRYLAMKEDGRLASKCVTDECFFERLESNNYNTYRSKYTSWYVALKR 120
Qy 407 TGQYKLGSKTGPQKAI 423
Db 121 TGQYKLGSKTGPQKAI 137
```

Search completed: May 3, 2004, 13:07:49
Job time : 16.9254 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:05:48 ; Search time 51.1726 Seconds
(without alignments)

2817.753 Million cell updates/sec

Title: US-09-775-964-5

Perfect score: 2398

Sequence: 1 PTDLRTNIGPDIMRVWAP.....LVTLPHNLPGLDVPST 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mmc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	824.5	34.4	196	4 P78443	P78443 homo sapien
2	762	31.8	153	11 Q925A3	Q925A3 mus musculus
3	760.5	31.7	2478	13 Q93406	Q93406 brachydanio
4	740.5	30.9	170	11 Q60487	Q60487 cavia porce
5	698	29.1	155	13 Q90Y92	Q90Y92 cynops pyr
6	676	28.2	130	6 Q77767	Q77767 canis faml
7	618	25.8	155	13 Q8QFR9	Q8qfr9 fugu rubrip
8	607.5	25.3	154	13 Q7ZKX5	Q7zdx5 brachydanio
9	585	24.4	111	6 Q9BDX1	Q9bdx1 macaca mula
10	561	23.4	108	6 Q9NLS7	Q9nls7 capreolus c
11	561	23.4	125	13 Q98TD8	Q98td8 cynops pyr
12	533	22.2	105	11 Q7TPG9	Q7tpg9 mus musculus
13	508	21.2	96	6 Q7YEN5	Q7yrn5 sus scrofa
14	484	20.2	109	11 Q925A1	Q925a1 mus musculus
15	480	20.0	112	11 Q925A2	Q925a2 mus musculus
16	476	19.8	101	13 P79706	P79706 cynops pyr

17	474.5	19.8	146	13	Q07659	Q07659 gallus gall
18	457	19.1	87	6	Q8WMP4	Q8wmp4 equus cabal
19	428.5	17.9	379	4	O95617	O95617 homo sapien
20	417	17.4	293	6	Q9XSG0	Q9xsg0 oryctolagus
21	415.5	17.3	810	11	Q8R3F3	Q8r3f3 mus musculus
22	396	16.5	1358	11	Q8BYI9	Q8byi9 mus musculus
23	394	16.4	1358	4	Q92752	Q92752 homo sapien
24	394	16.4	1358	4	Q15568	Q15568 homo sapien
25	383.5	16.0	354	4	Q9UQ56	Q9uq56 homo sapien
26	378.5	15.8	1356	11	Q05546	Q05546 rattus norv
27	358	14.9	1350	13	Q7T3T6	Q7t3t6 brachydanio
28	358	14.9	1353	13	Q00546	Q00546 gallus gall
29	341	14.2	76	6	Q9NOV2	Q9nov2 ovis aries
30	336	14.0	1532	13	Q90994	Q90994 gallus gall
31	332.5	13.9	114	4	Q16443	Q16443 homo sapien
32	332.5	13.9	114	4	O00527	O00527 homo sapien
33	329	13.7	1810	13	Q90824	Q90824 gallus gall
34	327	13.6	2019	11	Q64706	Q64706 mus musculus
35	327	13.6	2019	11	Q80YX2	Q80yx2 mus musculus
36	327	13.6	2110	11	Q80YX1	Q80yx1 mus musculus
37	326.5	13.6	614	13	Q90484	Q90484 brachydanio
38	318	13.3	619	11	Q62701	Q62701 rattus norv
39	317	13.2	1714	13	Q90995	Q90995 gallus gall
40	312.5	13.0	922	13	Q93405	Q93405 brachydanio
41	312	13.0	920	11	Q80U14	Q80ui4 mus musculus
42	307.5	12.8	215	13	Q7T0K8	Q7t0k8 ambystoma m
43	306.5	12.8	347	6	Q95KV5	Q95kv5 bos taurus
44	305	12.7	68	6	Q28692	Q28692 oryctolagus
45	302.5	12.6	4135	6	O18977	O18977 bos taurus

ALIGNMENTS

RESULT 1

P78443 ID P78443 PRELIMINARY; PRT; 196 AA.
AC P78443;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last amotation update)
DE 21 kDa basic fibroblast growth factor (BFGF).
GN BFGF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89184522; PubMed=2538817;
RA Prats H., Kaghad M., Prats A.C., Klagsbrun M., Lelias J.M.,
RA Liauun P., Chalon P., Tauber J.P., Analtic F., Smith J.A., Caput D.;
RT "High molecular mass forms of basic fibroblast growth factor are
RT initiated by alternative CUG codons.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:1836-1840(1989).
RN [2]
RP SEQUENCE OF 81-168 FROM N.A.
RX MEDLINE=93038590; PubMed=1417798;
RA Watson R., Anthony F., Pickett M., Lambden P., Masson G.M.,
RA Thomas E.J.;
RT "Reverse transcription with nested polymerase chain reaction shows
RT expression of basic fibroblast growth factor transcripts in human
RT granulosa and cumulus cells from in vitro fertilisation patients.";
RL Biochem. Biophys. Res. Commun. 187:1227-1231(1992).
DR EMBL; J04513; AA5252.1; -;
DR EMBL; S47380; AAD13853.1; -;
DR HSSP; P09038; 1BFF.
DR GO; GO:0008083; F:Growth factor activity; IEA.
DR InterPro; IPR008996; Cytok IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR ProDom; PD000831; IL1_HBGF; 1.


```
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Basic fibroblast growth factor (Fragment).
OS BFGF.
GN Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RA Trochta O.A., Jacobs R.M., LaMarre J.;
RT "The role bFGF in canine Hemangiosarcoma.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF060562; AAC35912.1; -.
DR GO; GO:0008083; F:growth factor activity; IEA.
DR InterPro; IPR008996; Cytok IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR ProDom; PD000831; IL1_HBGF; 1.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 14902 MW; 21900876E878FAEA CRC64;
Query Match 28.2%; Score 676; DB 6; Length 130;
Best Local Similarity 96.9%; Pred. No. 4.2e-44;
Matches 126; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 303 EKDPKRLKCKNGGFFLRIHPDGRVDGVRKSPDHKIQLOAERGVSIGVCANRYLAM 362
DB 1 FKPDKRLKCKNGGFFLRIHPDGRVDGVRKSPDHKIQLOAERGVSIGVCANRYLAM 60
QY 363 KEDGRLLASKCVTDECFERLESNNYTYRSRKYTSWYVALKRTGQYKLGSKTGPQOKA 422
DB 61 KEDGRLLASKCVTDECFERLESNNYTYRSRKYTSWYVALKRTGQYKLGSKTGPQOKA 120
QY 423 ILFLPMSAAS 432
DB 121 ILFLPMSAKS 130
RESULT 7
Q8QFR9 PRELIMINARY; PRT; 155 AA.
ID Q8QFR9
AC Q8QFR9;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Basic fibroblast growth factor.
GN FGF2.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RC Botcherby M.R.;
RT "Comparative vertebrate genomic sequence analysis studies based on
RT Fugu rubripes.";
RL Thesis (2001); University College London, London, United Kingdom.
DR EMBL; AJ426040; CAD19830.1; -.
DR GO; GO:0008083; F:growth factor activity; IEA.
DR InterPro; IPR008996; Cytok IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR ProDom; PD000831; IL1_HBGF; 1.
DT 01-JUN-2001 (TremBLrel. 17, Created)
ID Q9BDX1 PRELIMINARY; PRT; 111 AA.
AC Q9BDX1;
DT 01-JUN-2001 (TremBLrel. 17, Created)
```

```
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
SQ SEQUENCE 155 AA; 17113 MW; AFE12DBDC78FBBE CRC64;
Query Match 25.8%; Score 618; DB 13; Length 155;
Best Local Similarity 77.1%; Pred. No. 1.5e-39;
Matches 118; Conservative 5; Mismatches 30; Indels 0; Gaps 0;
QY 278 MAAGSIITLPALPEDGGGAPPGHFKDPKRLKCKNGGFFLRIHPDGRVDGVRKSDPHI 337
DB 1 MATGGIITLPSTPEDGGGGPPGSGFKDPKRLKCKNGGFFLRIHSDGAVDGTREKTDHI 60
QY 338 KLOQAERGVSIGVCANRYLAMKEDGRLLASKCVTDECFERLESNNYTYRSKY 397
DB 61 KLOQAATSVGEVWIKVCANRYLAMNRDGRFGMKRATDECHFLERLESNNYTYRSKY 120
QY 398 TSWYVALKRTGQYKLGSKTGPQOKAILFLPMSA 430
DB 121 PNMFVGLTRTGNYSKSGTKTGPCQKAILFLPMSA 153
RESULT 8
Q7ZKZ5 PRELIMINARY; PRT; 154 AA.
ID Q7ZKZ5
AC Q7ZKZ5;
DT 01-JUN-2003 (TremBLrel. 24, Created)
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Fibroblast growth factor 2.
GN FGF2.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC Groth C., Lardelli M.;
RT "Expression analysis of zebrafish fgf2 during embryogenesis.";
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY269790; AAP32155.1; -.
DR GO; GO:0008083; F:growth factor activity; IEA.
DR InterPro; IPR008996; Cytok IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR ProDom; PD000831; IL1_HBGF; 1.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
SQ SEQUENCE 154 AA; 17114 MW; 2B28A02B35E1B39B CRC64;
Query Match 25.3%; Score 607.5; DB 13; Length 154;
Best Local Similarity 75.2%; Pred. No. 9.4e-39;
Matches 115; Conservative 10; Mismatches 27; Indels 1; Gaps 1;
QY 278 MAAGSIITLPALPEDGGGAPPGHFKDPKRLKCKNGGFFLRIHPDGRVDGVRKSDPHI 337
DB 1 MATGGIITLPAP-DAENSFPAGSFRDPKRLKCKNGGFFLRINADGVDGARDKSDPHI 59
QY 338 KLOQAERGVSIGVCANRYLAMKEDGRLLASKCVTDECFERLESNNYTYRSKY 397
DB 60 RLQLOATAVGEVLKIGICTNRFAMNADGRFGTKRTTDECYFLERLESNNYTYRSKY 119
QY 398 TSWYVALKRTGQYKLGSKTGPQOKAILFLPMSA 430
DB 120 PDWYVALKRTGQYKSGSKTSFGQKAILFLPMSA 152
RESULT 9
Q9BDX1 PRELIMINARY; PRT; 111 AA.
ID Q9BDX1
AC Q9BDX1;
DT 01-JUN-2001 (TremBLrel. 17, Created)
```


DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Basic fibroblast growth factor (Fragment).
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
[1]
RN SEQUENCE FROM N.A.
RP Sekhon H.S., Keller J.K., Spindel E.R.;
RA "Alterations in Collagen and Elastin Gene Expression in Fetal
RT Pulmonary Vessels in Monkeys Following Prenatal Nicotine Exposure: A
RT Possible Role of alpha7 Nicotinic Acetylcholine Receptor in Persistent
RT Pulmonary Hypertension."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF251270; AAK37962.1; -.
DR HSSP; P09038; 2FGF.
DR GO; GO:0008083; F: growth factor activity; IEA.
DR InterPro; IPR008996; Cytok_IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR ProDom; PD000831; IL1_HBGF; 1.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
FT NON_TER 1 1
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12633 MW; EC0967A5261F5487 CRC64;

Query Match 24.4%; Score 585; DB 6; Length 111;
Best Local Similarity 100.0%; Pred. No. 3e-37;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 320 IHPDGRVDGVRKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTDEC 379
DB 1 IHPDGRVDGVRKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTDEC 60

QY 380 FFERLESNNYTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPMSA 430
DB 61 FFERLESNNYTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPMSA 111

RESULT 10
Q9N1S7 PRELIMINARY; PRT; 108 AA.
AC Q9N1S7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Basic fibroblast growth factor (Fragment).
GN BFGF.
OS Capreolus capreolus (Roe deer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervioidea;
OC Cervidae; Odocoileinae; Capreolus.
OX NCBI_TaxID=9858;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Testis;
RX MEDLINE=20532861; PubMed=11078967;
RA Wagener A., Blottner S., Goritz F., Fickel J.;
RT "Detection of growth factors in the testis of roe deer (Capreolus
RT capreolus).";
RL Anim. Reprod. Sci. 64:65-75(2000).
DR EMBL; AF152587; AAF73226.1; -.
DR HSSP; P09038; 4FGF.
DR GO; GO:0008083; F: growth factor activity; IEA.
DR InterPro; IPR008996; Cytok_IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR ProDom; PD000831; IL1_HBGF; 1.

DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
FT NON_TER 1 1
FT NON_TER 108 108
SQ SEQUENCE 108 AA; 12399 MW; 6BC7B7244214567E CRC64;

Query Match 23.4%; Score 561; DB 6; Length 108;
Best Local Similarity 98.1%; Pred. No. 2e-35;
Matches 106; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 319 RIHPDGRVDGVRKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTDEC 378
DB 1 RIHPDGRVDGVRKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTDEC 60

QY 379 FFERLESNNYTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFL 426
DB 61 FFERLESNNYTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFL 108

RESULT 11
Q98TD8 PRELIMINARY; PRT; 125 AA.
AC Q98TD8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fibroblast growth factor-2 (Fragment).
GN FGF-2.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
[1]
RN SEQUENCE FROM N.A.
RA Mizuno N., Hayashi T., Kondoh H., Okamoto M.;
RT "Cynops fibroblast growth factor-2";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB049625; BAB40835.1; -.
DR HSSP; P09038; 1BFF.
DR GO; GO:0008083; F: growth factor activity; IEA.
DR InterPro; IPR008996; Cytok_IL1_like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR ProDom; PD000831; IL1_HBGF; 1.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
FT NON_TER 1 1
SQ SEQUENCE 125 AA; 14244 MW; 5C27F41DC6E60C13 CRC64;

Query Match 23.4%; Score 561; DB 13; Length 125;
Best Local Similarity 86.3%; Pred. No. 2.5e-35;
Matches 107; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 309 LYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLL 368
DB 2 LYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLL 61

QY 369 LASKCVTDECFFERLESNNYTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPM 428
DB 62 MALKWITDECFFERLESNNYTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPM 121

QY 429 SAAS 432
DB 122 SAKS 125

RESULT 12
Q77PG9 PRELIMINARY; PRT; 105 AA.
ID Q77PG9
AC Q77PG9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DE	FGF2 (Fragment).	DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
GN	FGF2.	DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
OS	Mus musculus (Mouse).	DE	Fibroblast growth factor 2.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	GN	FGF2.
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	OS	Mus musculus (Mouse).
OX	NCBI_TaxID=10090;	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
RN	[1]	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RP	SEQUENCE FROM N.A.	OX	NCBI_TaxID=10090;
RC	STRAIN=129/SVJ;	RN	[1]
RA	Foletti A., Vuadens F., Beermann F.;	RP	SEQUENCE FROM N.A.
RL	"Intracellular localization of mouse FGF2.";	RC	STRAIN=FVB/N;
RL	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.	RA	Dirks R.P., Griep A.E.;
DR	EMBL; AY324451; AAP92385.1; JOINED.	RT	"Multiple novel variants of fibroblast growth factor 2 transcripts are expressed in mouse embryos.";
DR	EMBL; AY324449; AAP92385.1; JOINED.	RT	expressed in mouse embryos.";
DR	EMBL; AY324450; AAP92385.1; JOINED.	RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
FT	NON_TER 1	DR	EMBL; AY027558; AAK52310.1; -
SQ	SEQUENCE 105 AA; 11945 MW; C24E39323A79D469 CRC64;	DR	GO; GO:0008083; F:growth factor activity; IEA.
		DR	InterPro; IPR008996; Cytok IL1-like.
		DR	InterPro; IPR002348; IL1_HBGF.
		DR	Pfam; PF00167; FGF; 1.
		DR	PRINTS; PR00262; IL1HBGF.
		DR	ProDom; PD000831; IL1_HBGF; 1.
		DR	SMART; SM00442; FGF; 1.
		DR	PROSITE; PS00247; HBGF_FGF; 1.
		SQ	SEQUENCE 109 AA; 12388 MW; 61074ADE3303C860 CRC64;
		Query Match	20.2%; Score 484; DB 11; Length 109;
		Best Local Similarity	96.9%; Pred. No. 1.6e-29;
		Matches	93; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY	328 GVREKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTDECFPERLESN 387		
Db	1 GVREKSDPHVQLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTECFPERLESN 60		
QY	388 NNTYRSRKYTSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432		
Db	61 NNTYRSRKYTSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 105		
RESULT 13			
Q7YRNS			
ID	Q7YRNS PRELIMINARY; PRT; 96 AA.		
AC	Q7YRNS;		
DT	01-OCT-2003 (TrEMBLrel. 25, Created)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Fibroblast growth factor 2 (Fragment).		
GN	FGF2.		
OS	Sus scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
OX	NCBI_TaxID=9823;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Endometrium;		
RA	Binspanier R., Welter H.;		
RT	"Growth factors during porcine endometrial development.";		
RL	Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ577089; CAEL1791.1; -		
FT	NON_TER 1		
FT	NON_TER 96		
SQ	SEQUENCE 96 AA; 11094 MW; F8CA875BED095FA3 CRC64;		
		Query Match	21.2%; Score 508; DB 6; Length 96;
		Best Local Similarity	99.0%; Pred. No. 1.9e-31;
		Matches	95; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY	313 NGGFFLRHPDGRVDGVREKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASK 372		
Db	1 NGGFFLRHPDGRVDGVREKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASK 60		
QY	373 CVTDECFPERLESNNYTRSRKYTSWYVALKRTG 408		
Db	61 CVTDECFPERLESNNYTRSRKYTSWYVALKRTG 96		
RESULT 14			
Q925A1			
ID	Q925A1 PRELIMINARY; PRT; 109 AA.		
AC	Q925A1;		
		Query Match	20.0%; Score 480; DB 11; Length 112;
		Best Local Similarity	96.8%; Pred. No. 3.3e-29;
		Matches	93; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY	337 IKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTDECFPERLESNNYTRSRK 396		
Db	14 IKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTECFPERLESNNYTRSRK 73		
QY	397 YTSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432		
Db	74 YSSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 109		
RESULT 15			
Q925A2			
ID	Q925A2 PRELIMINARY; PRT; 112 AA.		
AC	Q925A2;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Fibroblast growth factor 2.		
GN	FGF2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FVB/N;		
RA	Dirks R.P., Griep A.E.;		
RT	"Multiple novel variants of fibroblast growth factor 2 transcripts are expressed in mouse embryos.";		
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY027557; AAK52309.1; -		
DR	GO; GO:0008083; F:growth factor activity; IEA.		
DR	InterPro; IPR008996; Cytok IL1-like.		
DR	InterPro; IPR002348; IL1_HBGF.		
DR	Pfam; PF00167; FGF; 1.		
DR	PRINTS; PR00262; IL1HBGF.		
DR	ProDom; PD000831; IL1_HBGF; 1.		
DR	SMART; SM00442; FGF; 1.		
DR	PROSITE; PS00247; HBGF_FGF; 1.		
SQ	SEQUENCE 112 AA; 12725 MW; B00557ABE0257CCB CRC64;		
		Query Match	20.0%; Score 480; DB 11; Length 112;
		Best Local Similarity	96.8%; Pred. No. 3.3e-29;
		Matches	93; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY	337 YTSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432		
Db	74 YSSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 109		

	Matches	92;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	338	KLQLQAEERGVS	IKGVCANRYLAMKEDGRLLASKCVTDECFFFERLESNNYNTYRSKY	397						
Db	18	KLQLQAEERGVS	IKGVCANRYLAMKEDGRLLASKCVTDECFFFERLESNNYNTYRSKY	77						
Qy	398	TSMYVALKRTGQYK	LGSKTGPQKAILFLPMSAKS	432						
Db	78	SSWYVALKRTGQYK	LGSKTGPQKAILFLPMSAKS	112						

Search completed: May 3, 2004, 13:11:07
Job time : 53.1726 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:05:13 ; Search time 29.5054 Seconds
(without alignments)
1781.157 Million cell updates/sec

Title: US-09-775-964-6
Perfect score: 1079
Sequence: 1 GIRLKGTKGKEDGFGPFG.....PPGPPKGLGHPGQRGET 186

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Genesecp1980s:*
- 2: Genesecp1990s:*
- 3: Genesecp2000s:*
- 4: Genesecp2001s:*
- 5: Genesecp2002s:*
- 6: Genesecp2003as:*
- 7: Genesecp2003bs:*
- 8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1079	100.0	186	2 AAR34587	Aar34587 Fibronect
2	1079	100.0	186	2 AAW33341	Aaw33341 Oligopept
3	1079	100.0	186	2 AAY05462	Aay05462 Fibronect
4	1079	100.0	464	2 AAR34589	Aar34589 Fibronect
5	1079	100.0	464	2 AAW33342	Aaw33342 Protein u
6	1079	100.0	464	2 AAY05461	Aay05461 Fibronect
7	1079	100.0	1838	2 AAR33257	Aar33257 Human col
8	1079	100.0	1838	7 ADES5566	Ades5566 Human Pro
9	1079	100.0	1838	7 ADES5570	Ades5570 Human Pro
10	1069	99.1	489	2 AAW33343	Aaw33343 Protein u
11	930	86.2	1806	5 AAB84266	Aab84266 Human end
12	930	86.2	1806	5 AABJ05596	Aabj05596 Breast ca
13	930	86.2	1806	6 ABR58545	Abr58545 Human can
14	930	86.2	1806	6 AABU56581	Aabu56581 Lung canc
15	835	77.4	454	2 AAR25155	Aar25155 Fibronect
16	834	77.3	167	2 AAR22273	Aar22273 Human col
17	741	68.7	1053	4 AAB85863	Aab85863 Murine ad
18	741	68.7	1739	4 AAG77792	Aag77792 Murine pr
19	737	68.3	1745	4 AAG77793	Aag77793 Human pro
20	737	68.3	1745	5 AAB97234	Aab97234 Novel hum
21	509.5	47.2	1466	4 AAE02534	Aae02534 Bovine al
22	509.5	47.2	1466	4 AAE02533	Aae02533 Bovine al
23	503.5	46.7	357	2 AAR95115	Aar95115 Interveni
24	503.5	46.7	357	2 AAW57646	Aaw57646 Collagen-
25	503.5	46.7	357	4 AAB64008	Aab64008 CLP-CB pr

26	503.5	46.7	357	4 AAB72738	Aab72738 Repetitiv
27	501.5	46.5	772	2 AAR23873	Aar23873 Human alp
28	501.5	46.5	772	2 AAW09643	Aaw09643 Human typ
29	501.5	46.5	1685	4 ABG04839	Abg04839 Novel hum
30	501.5	46.5	1693	4 ABG15619	Abg15619 Novel hum
31	501	46.4	1669	5 ABB57334	Abb57334 Mouse isc
32	495.5	45.9	1064	2 AAR93254	Aar93254 Collagen-
33	495.5	45.9	1064	2 AAW57652	Aaw57652 Collagen-
34	495.5	45.9	1065	2 AAR37741	Aar37741 Collagen-
35	494	45.8	1606	6 ABR40002	Abr40002 Human col
36	492.5	45.6	1466	4 AAE02537	Aae02537 Porcine a
37	492	45.6	654	4 AAG63332	Aag63332 Anino aci
38	491	45.5	1496	5 ABR93944	Abr93944 Human pol
39	491	45.5	1496	6 ABR47419	Abr47419 Breast ca
40	491	45.5	1496	7 ADB70382	Adb70382 Procollag
41	491	45.5	1496	7 ADE61180	Ade61180 Human Pro
42	488	45.2	330	2 AAW57645	Aaw57645 Collagen-
43	488	45.2	408	2 AAW07539	Aaw07539 Collagen
44	488	45.2	408	4 AAB64007	Aab64007 CLP prote
45	488	45.2	408	4 AAB72737	Aab72737 Repetitiv

ALIGNMENTS

RESULT 1
AAR34587
ID AAR34587 standard; protein; 186 AA.

XX AC AAR34587;

XX XX

DT 14-SEP-1993 (first entry)

XX XX

DE DE

XX XX

KW Polymerase chain reaction; amplify; human; primer; h-ColV; pTYCOLV;

KW intracellular adhesion; PCR; domain; fibronectin; insulin binding; ss.

XX OS Synthetic.

XX XX

PN JP05097698-A.

XX XX

PD 20-APR-1993.

XX XX

PF 14-OCT-1991; 91JP-00291958.

XX XX

PR 14-OCT-1991; 91JP-00291958.

XX XX

PA (TAKI) TAKARA SHUZO CO LTD.

DR WPI; 1993-164369/20.

XX XX

PT Artificial functional polypeptide having insulin-combining activity -

PT includes inter-cellular adhesive domain polypeptide of human fibronectin

PT and polypeptide contg. specific aminoacid sequence.

XX XX

PS Disclosure; Page 7-8; 12pp; Japanese.

XX XX

CC The sequences given in AAR34585-90 are polypeptides which were used in

CC the production of recombinant polypeptides which contained the

CC intracellular adhesive domain of human fibronectin, a spacer and an

CC insulin binding domain. These polypeptides have insulin-binding and

CC intercellular adhesion activity

XX XX

SQ Sequence 186 AA;

Query Match 100.0%; Score 1079; DB 2; Length 186;

Best Local Similarity 100.0%; Pred. No. 3.7e-69;

Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGFGPFGKDMGKIDRGKIDGPPRGEDGPEGPKRGGNGPGPL 60

DB 1 GIRLKGTKGKEDGFGPFGKDMGKIDRGKIDGPPRGEDGPEGPKRGGNGPGPL 60

```
QY 61 GPPGKGLVPLPGYPRQPKSGIGFPGPGANGKGGRTGKPGPRGQRGTGPR 120
DB 61 GPPGKGLVPLPGYPRQPKSGIGFPGPGANGKGGRTGKPGPRGQRGTGPR 120

QY 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180
DB 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180

QY 181 GORGET 186
DB 181 GORGET 186

RESULT 2
AAW33341
ID AAW33341 standard; protein; 186 AA.
XX AC AAW33341;
XX DT 23-FEB-1998 (first entry)
XX DE Oligopeptide ColV.
XX KW Oligopeptide ColV; target cell; transfection; retroviral vector;
XX KW gene therapy; cancer; viral disease; acquired immunodeficiency syndrome;
XX KW AIDS.
XX OS Synthetic.
XX PN WO9718318-A1.
XX PD 22-MAY-1997.
XX PF 07-NOV-1996; 96WO-JP003254.
XX PR 13-NOV-1995; 95JP-00294382.
XX PR 08-MAR-1996; 96JP-00051847.
XX PA (TAKI ) TAKARA SHUZO CO LTD.
XX PI Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;
XX DR WPI; 1997-289294/26.
XX PT Method for increasing efficacy of gene transfer to target cell using
XX PT retrovirus - by infection of the target cell in the presence of a
XX PT substance which binds to the virus and a substance which binds to the
XX PT target cell.
XX PS Claim 44; Page 100-101; 194pp; Japanese.
XX CC The present sequence is the oligopeptide ColV, which was used in the
XX CC development of a novel method for increasing the efficiency of gene
XX CC introduction into a target cell using a retroviral vector. The method
XX CC comprises carrying out viral infection of the target cell in the presence
XX CC of a retrovirus and target cell binding substance or substances. The
XX CC method can be used to effectively introduce genes into target cells for
XX CC the gene therapy of cancer and viral diseases, e.g. AIDS.
XX SQ Sequence 186 AA;
Query Match 100.0%; Score 1079; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.7e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRLKGTKGKGGEDGFPFGKDMGDKDRGEIGPPGPRGEDGPEGKRGGPNGDGPL 60
DB 1 GIRLKGTKGKGGEDGFPFGKDMGDKDRGEIGPPGPRGEDGPEGKRGGPNGDGPL 60
QY 61 GPPGKGLVPLPGYPRQPKSGIGFPGPGANGKGGRTGKPGPRGQRGTGPR 120
DB 61 GPPGKGLVPLPGYPRQPKSGIGFPGPGANGKGGRTGKPGPRGQRGTGPR 120
QY 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180
DB 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180
QY 181 GORGET 186
DB 181 GORGET 186

RESULT 3
AAW33341
ID AAW33341 standard; protein; 186 AA.
XX AC AAW33341;
XX DT 07-JUL-1999 (first entry)
XX DE Fibronectin receptor targetting HIV strain ColV.
XX KW Fibronectin receptor; HIV; infection; therapy.
XX OS Unidentified.
XX PN JP10029952-A.
XX PD 03-FEB-1998.
XX PF 16-JUL-1996; 96JP-00185893.
XX PR 16-JUL-1996; 96JP-00185893.
XX PA (TAKI ) TAKARA SHUZO CO LTD.
XX DR WPI; 1998-163674/15.
XX PT Control of human immunodeficiency virus infection - using composition
XX PT comprising replication defective HIV vector.
XX PS Example 5; Page 23-24; 24pp; Japanese.
XX CC This sequence represents a fibronectin receptor that can be used in the
XX CC method of the invention. The method is for the control of human
XX CC immunodeficiency virus (HIV) infection using a composition which
XX CC comprises a functional substance which participates in the infection of
XX CC HIV. The method is used to control HIV-infection.
XX SQ Sequence 186 AA;
Query Match 100.0%; Score 1079; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.7e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRLKGTKGKGGEDGFPFGKDMGDKDRGEIGPPGPRGEDGPEGKRGGPNGDGPL 60
DB 1 GIRLKGTKGKGGEDGFPFGKDMGDKDRGEIGPPGPRGEDGPEGKRGGPNGDGPL 60
QY 61 GPPGKGLVPLPGYPRQPKSGIGFPGPGANGKGGRTGKPGPRGQRGTGPR 120
DB 61 GPPGKGLVPLPGYPRQPKSGIGFPGPGANGKGGRTGKPGPRGQRGTGPR 120
QY 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180
DB 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180
QY 181 GORGET 186
DB 181 GORGET 186

RESULT 4
AAR34589
ID AAR34589 standard; protein; 464 AA.
```

```
QY 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180
DB 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180

QY 181 GORGET 186
DB 181 GORGET 186

RESULT 3
AAW33341
ID AAW33341 standard; protein; 186 AA.
XX AC AAW33341;
XX DT 07-JUL-1999 (first entry)
XX DE Fibronectin receptor targetting HIV strain ColV.
XX KW Fibronectin receptor; HIV; infection; therapy.
XX OS Unidentified.
XX PN JP10029952-A.
XX PD 03-FEB-1998.
XX PF 16-JUL-1996; 96JP-00185893.
XX PR 16-JUL-1996; 96JP-00185893.
XX PA (TAKI ) TAKARA SHUZO CO LTD.
XX DR WPI; 1998-163674/15.
XX PT Control of human immunodeficiency virus infection - using composition
XX PT comprising replication defective HIV vector.
XX PS Example 5; Page 23-24; 24pp; Japanese.
XX CC This sequence represents a fibronectin receptor that can be used in the
XX CC method of the invention. The method is for the control of human
XX CC immunodeficiency virus (HIV) infection using a composition which
XX CC comprises a functional substance which participates in the infection of
XX CC HIV. The method is used to control HIV-infection.
XX SQ Sequence 186 AA;
Query Match 100.0%; Score 1079; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.7e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRLKGTKGKGGEDGFPFGKDMGDKDRGEIGPPGPRGEDGPEGKRGGPNGDGPL 60
DB 1 GIRLKGTKGKGGEDGFPFGKDMGDKDRGEIGPPGPRGEDGPEGKRGGPNGDGPL 60
QY 61 GPPGKGLVPLPGYPRQPKSGIGFPGPGANGKGGRTGKPGPRGQRGTGPR 120
DB 61 GPPGKGLVPLPGYPRQPKSGIGFPGPGANGKGGRTGKPGPRGQRGTGPR 120
QY 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180
DB 121 GERGRGITGKPGKNSGGDPAGPPGPRGPNQPGTGFPGKGGPKDGLPGHP 180
QY 181 GORGET 186
DB 181 GORGET 186

RESULT 4
AAR34589
ID AAR34589 standard; protein; 464 AA.
```

XX AAR34589;
XX
XX 14-SEP-1993 (first entry)
XX
XX Fibronection domain #3.
XX
XX Polymerase chain reaction; amplify; human; primer; h-ColV; pTVColV;
KW intracellular adhesion; PCR; domain; fibronection; insulin binding; es.
KW
XX Synthetic.
XX
XX JF05097698-A.
XX
XX 20-APR-1993.
XX
XX 14-OCT-1991; 91JP-00291958.
XX
XX 14-OCT-1991; 91JP-00291958.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX WPI; 1993-164369/20.
XX
XX Artificial functional polypeptide having insulin-combining activity -
PT includes inter-cellular adhesive domain polypeptide of human fibronection
PT and polypeptide contg. specific aminoacid sequence.
XX
XX Disclosure; Page 9-10; 12pp; Japanese.
XX
XX The sequences given in AAR34585-90 are polypeptides which were used in
CC the production of recombinant polypeptides which contained the
CC intracellular adhesive domain of human fibronection, a spacer and an
CC insulin binding domain. These polypeptides have insulin-binding and
CC intercellular adhesion activity
XX
SQ Sequence 464 AA;
Query Match 100.0%; Score 1079; DB 2; Length 464;
Best Local Similarity 100.0%; Pred. No. 7.9e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRGLKGTGKEGEDGPPGFKGDMGKGDREIGTGPGRGEGDPEGPKGRGGNGDPGPL 60
Db 279 GIRGLKGTGKEGEDGPPGFKGDMGKGDREIGTGPGRGEGDPEGPKGRGGNGDPGPL 338
QY 61 GPPEKGLGVPLPGYPRGQPKSGTGFPGFGANGKEKGRCTPKGPRGORGPTGPR 120
Db 339 GPPEKGLGVPLPGYPRGQPKSGTGFPGFGANGKEKGRCTPKGPRGORGPTGPR 398
QY 121 GERGPRGITGKPGKNSGSDGPPAGPPGERGPNPQGTGFPKGPDPGKDLPGHP 180
Db 399 GERGPRGITGKPGKNSGSDGPPAGPPGERGPNPQGTGFPKGPDPGKDLPGHP 458
QY 181 GORGET 186
Db 459 GORGET 464
RESULT 5
AAW33342
ID AAW33342 standard; protein; 464 AA.
XX
XX AAW33342;
XX
XX 23-FEB-1998 (first entry)
XX
XX Protein used in development of gene transfer method.
XX
XX Target cell; transfection; retroviral vector; gene therapy; cancer;
KW viral disease; acquired immunodeficiency syndrome; AIDS.
XX
XX Synthetic.

XX WO9718318-A1.
XX
XX 22-MAY-1997.
XX
XX 07-NOV-1996; 96WO-JP003254.
XX
XX 13-NOV-1995; 95JP-00294382.
PR 08-MAR-1996; 96JP-00051847.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;
PI WPI; 1997-289294/26.
XX
XX Method for increasing efficacy of gene transfer to target cell using
PT retrovirus - by infection of the target cell in the presence of a
PT substance which binds to the virus and a substance which binds to the
PT target cell.
XX
XX Claim 45; Page 100-104; 194pp; Japanese.
XX
XX The present sequence was used in the development of a novel method for
CC increasing the efficiency of gene introduction into a target cell using a
CC retroviral vector. The method comprises carrying out viral infection of
CC the target cell in the presence of a retrovirus and target cell binding
CC substance or substances. The method can be used to effectively introduce
CC genes into target cells for the gene therapy of cancer and viral
CC diseases, e.g. AIDS
XX
SQ Sequence 464 AA;
Query Match 100.0%; Score 1079; DB 2; Length 464;
Best Local Similarity 100.0%; Pred. No. 7.9e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRGLKGTGKEGEDGPPGFKGDMGKGDREIGTGPGRGEGDPEGPKGRGGNGDPGPL 60
Db 279 GIRGLKGTGKEGEDGPPGFKGDMGKGDREIGTGPGRGEGDPEGPKGRGGNGDPGPL 338
QY 61 GPPEKGLGVPLPGYPRGQPKSGTGFPGFGANGKEKGRCTPKGPRGORGPTGPR 120
Db 339 GPPEKGLGVPLPGYPRGQPKSGTGFPGFGANGKEKGRCTPKGPRGORGPTGPR 398
QY 121 GERGPRGITGKPGKNSGSDGPPAGPPGERGPNPQGTGFPKGPDPGKDLPGHP 180
Db 399 GERGPRGITGKPGKNSGSDGPPAGPPGERGPNPQGTGFPKGPDPGKDLPGHP 458
QY 181 GORGET 186
Db 459 GORGET 464
RESULT 6
AAY05461
ID AAY05461 standard; protein; 464 AA.
XX
XX AAY05461;
XX
XX 07-JUL-1999 (first entry)
XX
XX Fibronection receptor targetting HIV strain C277-Colv.
XX
XX Fibronection receptor; HIV; infection; therapy.
XX
XX Unidentified.
XX
XX JP10029952-A.
XX
XX 03-FEB-1998.
XX
XX 16-JUL-1996; 96JP-00185893.
XX
XX

```
XX 16-JUL-1996; 96JP-00185893.
XX (TAKI ) TAKARA SHUZO CO LTD.
XX WPI; 1998-163674/15.
XX Control of human immunodeficiency virus infection - using composition
XX comprising replication defective HIV vector.
XX Example 4; Page 22-23; 24pp; Japanese.
XX This sequence represents a fibronectin receptor that can be used in the
XX method of the invention. The method is for the control of human
XX immunodeficiency virus (HIV) infection using a composition which
XX comprises a functional substance which participates in the infection of
XX HIV. The method is used to control HIV-infection
XX Sequence 464 AA;
Query Match 100.0%; Score 1079; DB 2; Length 464;
Best Local Similarity 100.0%; Pred. No. 7.9e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRGLKGTGKXGEGDGFPGFKGDMGIKDRGEIGPPRGEDGPGPKRGPGNDPGPL 60
Db 279 GIRGLKGTGKXGEGDGFPGFKGDMGIKDRGEIGPPRGEDGPGPKRGPGNDPGPL 338
QY 61 GPPGEKGLGVPLPGYPRQPKSGIGFPFPGANGKGRGTGPKPGRQGRGTGPR 120
Db 339 GPPGEKGLGVPLPGYPRQPKSGIGFPFPGANGKGRGTGPKPGRQGRGTGPR 398
QY 121 GERGRGITGKPGKNSGDDGPPGPRGNGPQGTGFPKGPDPGPKDGLPGHP 180
Db 399 GERGRGITGKPGKNSGDDGPPGPRGNGPQGTGFPKGPDPGPKDGLPGHP 458
QY 181 GORGET 186
Db 459 GORGET 464
RESULT 7
AAR53257
ID AAR53257 standard; protein; 1838 AA.
XX AAR53257;
XX 12-JAN-1995 (first entry)
XX Human collagen (Type V).
XX Human collagen; alpha 1; V type collagen; placental mRNA.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..38
XX /label= signal_peptide
XX Region 444..538
XX /note= "contains (Gly-X-Y) repeats"
XX Binding-site 645..647
XX /label= RGD
XX /note= "cell adhesion motif"
XX Binding-site 663..665
XX /label= RGD
XX /note= "cell adhesion motif"
XX Domain 897..933
XX /label= heparin_binding_domain
XX Region 1573..1838
XX /label= C-terminal_region
XX /note= "contains 8 Cys residues"
XX JP06105687-A.
```

```
XX 19-APR-1994.
XX 27-DEC-1991; 91JP-00358300.
XX 27-DEC-1991; 91JP-00358300.
XX (TAKI ) TAKARA SHUZO CO LTD.
XX WPI; 1994-163129/20.
XX N-PSDB; AAQ64556.
XX Human collagen V-type gene - is used for diagnosis of human collagen V-
XX type related diseases.
XX Claim 1; Page 6-14; 19pp; Japanese.
XX This amino acid sequence of type V collagen contains several distinct
XX domains including a region comprising repeated (Gly-X-Y) motifs and a
XX central domain containing two copies of the Arg-Gly-Asp cell adhesion
XX motif. The cDNA sequence encoding type V collagen was isolated from a
XX human placental library and will be useful for diagnosis of diseases
XX related to type V collagen
XX Sequence 1838 AA;
Query Match 100.0%; Score 1079; DB 2; Length 1838;
Best Local Similarity 100.0%; Pred. No. 2.6e-68;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRGLKGTGKXGEGDGFPGFKGDMGIKDRGEIGPPRGEDGPGPKRGPGNDPGPL 60
Db 799 GIRGLKGTGKXGEGDGFPGFKGDMGIKDRGEIGPPRGEDGPGPKRGPGNDPGPL 858
QY 61 GPPGEKGLGVPLPGYPRQPKSGIGFPFPGANGKGRGTGPKPGRQGRGTGPR 120
Db 859 GPPGEKGLGVPLPGYPRQPKSGIGFPFPGANGKGRGTGPKPGRQGRGTGPR 918
QY 121 GERGRGITGKPGKNSGDDGPPGPRGNGPQGTGFPKGPDPGPKDGLPGHP 180
Db 919 GERGRGITGKPGKNSGDDGPPGPRGNGPQGTGFPKGPDPGPKDGLPGHP 978
QY 181 GORGET 186
Db 979 GORGET 984
RESULT 8
ADE55566
ID ADE55566 standard; protein; 1838 AA.
XX ADE55566;
XX 29-JAN-2004 (first entry)
XX Human Protein P20908, SEQ ID NO 1383.
XX Human; pain; neuronal tissue; gene therapy;
XX spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX Homo sapiens.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
XX 01-NOV-2001; 2001US-0346382P.
XX 26-NOV-2001; 2001US-0333347P.
XX
```


Db 919 GERGPRGITGKPGKNSGGDGAGPFGGERGNGPQGTGFGPKGPPGPKGGLPGHP 978
181 GORGET 186
979 GORGET 984
RESULT 10
AAW33343
ID AAW33343 standard; protein; 489 AA.
XX AAW33343;
XX 23-FEB-1998 (first entry)
XX Protein used in development of gene transfer method.
XX Target cell; transfection; retroviral vector; gene therapy; cancer;
KW viral disease; acquired immunodeficiency syndrome; AIDS.
XX Synthetic.
XX WO9718318-Al.
XX 22-MAY-1997.
XX 07-NOV-1996; 96WO-JP003254.
XX 13-NOV-1995; 95JP-00294382.
PR 08-MAR-1996; 96JP-00051847.
XX (TAKI) TAKARA SHUZO CO LTD.
XX Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;
XX WPI; 1997-289294/26.
XX Method for increasing efficacy of gene transfer to target cell using
PT retrovirus - by infection of the target cell in the presence of a
PT substance which binds to the virus and a substance which binds to the
PT target cell.
XX Claim 45; Page 104-107; 194pp; Japanese.
XX The present sequence was used in the development of a novel method for
CC increasing the efficiency of gene introduction into a target cell using a
CC retroviral vector. The method comprises carrying out viral infection of
CC the target cell in the presence of a retrovirus and target cell binding
CC substance or substances. The method can be used to effectively introduce
CC genes into target cells for the gene therapy of cancer and viral
CC diseases, e.g. AIDS
XX Sequence 489 AA;
SQ Query Match 99.1%; Score 1069; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 4.2e-68;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRLKGITGKEGEDGFFGFKGDMGIKGRGEIGPPRGGEDGPEGPKRGSGPNDPGPL 60
279 GIRLKGITGKEGEDGFFGFKGDMGIKGRGEIGPPRGGEDGPEGPKRGSGPNDPGPL 338
QY 61 GPPGEGKGLGVPLGYPGRQGPKGSGIGFPFGPANGKGGRTGPKGPRGORGTPGR 120
339 GPPGEGKGLGVPLGYPGRQGPKGSGIGFPFGPANGKGGRTGPKGPRGORGTPGR 398
QY 121 GERGPRGITGKPGKNSGGDGAGPFGGERGNGPQGTGFGPKGPPGPKGGLPGHP 180
399 GERGPRGITGKPGKNSGGDGAGPFGGERGNGPQGTGFGPKGPPGPKGGLPGHP 458
QY 181 GORG 184
|||

Db 459 GORG 462
RESULT 11
AAU84266
ID AAU84266 standard; protein; 1806 AA.
XX AAU84266;
XX 08-MAY-2002 (first entry)
XX Human endometrial cancer related protein, COL11A1.
XX Human; endometrial cancer; differential expression; DNA microarray;
KW protein microarray.
XX Homo sapiens.
XX WO200209573-A2.
PN 07-FEB-2002.
PD 31-JUL-2001; 2001WO-US024104.
PF 31-JUL-2000; 2000US-0221735P.
PR (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX Mutter GL;
XX WPI; 2002-179967/23.
DR N-PSDB; ABK35486.
XX Diagnosing endometrial cancer comprises determining expression of nucleic
PT acid molecules or expression products that are differentially expressed
PT in normal and malignant endometrium.
XX Claim 33; Page 142-147; 233pp; English.
XX The invention relates to diagnosing endometrial cancer in a subject
CC suspected of having endometrial cancer comprising determining the
CC expression of a set of nucleic acid molecules or expression products in
CC an endometrial sample suspected of being cancerous, where the set of
CC nucleic acid molecules comprises at least 2 nucleic acid molecules
CC selected from 50 fully defined sequences as given in the specification.
CC The nucleic acids are used as an array of at least 2 of the 50 nucleic
CC acids bound to a solid substrate. Also included is a solid-phase protein
CC microarray comprising at least 2 antibodies or its antigen binding
CC fragments, that specifically bind at least 2 different polypeptides from
CC the 50 fully defined sequences as given in the specification, fixed to a
CC solid substrate. The methods and arrays are useful for the diagnosis of
CC endometrial cancer, selecting and monitoring treatment regimes and
CC identification of lead compounds useful for the treatment of endometrial
CC cancer. The present sequence is one of 50 proteins differentially
CC expressed between cancerous and non-cancerous samples
XX Sequence 1806 AA;
SQ Query Match 86.2%; Score 930; DB 5; Length 1806;
Best Local Similarity 86.0%; Pred. No. 8.8e-58;
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;
QY 1 GIRLKGITGKEGEDGFFGFKGDMGIKGRGEIGPPRGGEDGPEGPKRGSGPNDPGPL 60
769 GVRGLKSGKSGKSGEDGFFGFKGDMGLKGRGEVGIQPRGXDPGPGPKGRAGTGDGPGPS 828
QY 61 GPPGEGKGLGVPLGYPGRQGPKGSGIGFPFGPANGKGGRTGPKGPRGORGTPGR 120
829 GQAGEKGLGVPLGYPGRQGPKGSGIGFPFGPANGKGGRTGPKGPRGORGTPGR 888
QY 121 GERGPRGITGKPGKNSGGDGAGPFGGERGNGPQGTGFGPKGPPGPKGGLPGHP 180
889 GSRGARGTGPGRKGTSGDGPDPGGERGPGQGVGFPKPGPPGPPGPPGPPGPPGPPGPP 948

[illegible]

QY

1 GIRLGLTKXGEGDGFPGFKDMGIKDRGEIGPPGRGEDGGPGRGSGPDGPPL 60
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
CC therapeutic targets, in particular, the nucleic acid is useful for
CC bladder, brain, breast, cervix, colon/rectum kidney, lung, ovary,

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:07:19 ; Search time 9.54588 Seconds
(without alignments)
1005.924 Million cell updates/sec

Title: US-09-775-964-6
Perfect score: 1079
Sequence: 1 GINGLGTGKEGDEGPGF.....PPGPGKGLPHGPGRGST 186

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1079	100.0	186	4	US-09-366-009-6
2	1079	100.0	186	4	US-08-809-156B-6
3	1079	100.0	464	2	US-08-836-854-19
4	1079	100.0	464	4	US-09-366-009-7
5	1079	100.0	464	4	US-08-809-156B-7
6	1069	99.1	489	4	US-09-366-009-8
7	1069	99.1	489	4	US-08-809-156B-8
8	834	77.3	446	2	US-08-836-854-15
9	503.5	46.7	357	1	US-07-609-716-66
10	503.5	46.7	357	1	US-08-642-255-33
11	503.5	46.7	357	3	US-08-475-411A-66
12	503.5	46.7	357	3	US-08-478-029A-66
13	495.5	45.9	1064	1	US-08-642-255-32
14	488	45.2	330	1	US-08-642-255-32
15	488	45.2	408	1	US-07-609-716-65
16	488	45.2	408	3	US-08-475-411A-65
17	488	45.2	408	3	US-08-478-029A-65
18	486.5	45.1	1057	3	US-08-931-820-4
19	486.5	45.1	1078	3	US-08-963-825-21
20	486.5	45.1	1078	4	US-09-500-811-21
21	486.5	45.1	1078	4	US-09-570-573-21
22	486.5	45.1	1078	4	US-09-548-608-21
23	486	45.0	1024	3	US-08-931-820-2
24	486	45.0	1366	3	US-08-963-825-19
25	486	45.0	1366	4	US-09-500-811-19
26	486	45.0	1366	4	US-09-570-573-19
27	486	45.0	1366	4	US-09-548-608-19

28	483.5	44.8	252	1	US-08-642-255-61	Sequence 61, Appl
29	483	44.8	492	4	US-08-468-996-12	Sequence 12, Appl
30	481.5	44.6	532	1	US-08-494-168-9	Sequence 9, Appl
31	479.5	44.4	471	2	US-08-399-889-24	Sequence 24, Appl
32	479.5	44.4	471	3	US-09-167-364-24	Sequence 24, Appl
33	479.5	44.4	471	3	US-09-439-897-2	Sequence 2, Appl
34	478.5	44.3	1017	4	US-08-468-996-10	Sequence 10, Appl
35	478.5	44.3	1057	3	US-08-931-820-1	Sequence 1, Appl
36	478.5	44.3	1341	3	US-08-963-825-18	Sequence 18, Appl
37	478.5	44.3	1341	4	US-09-500-811-18	Sequence 18, Appl
38	478.5	44.3	1341	4	US-09-570-573-18	Sequence 18, Appl
39	478.5	44.3	1341	4	US-09-548-608-18	Sequence 18, Appl
40	478.5	44.3	1461	4	US-09-585-887-9	Sequence 9, Appl
41	478.5	44.3	1461	4	US-09-289-578-9	Sequence 9, Appl
42	478.5	44.3	1366	4	US-09-331-347C-21	Sequence 21, Appl
43	478	44.3	1366	4	US-09-585-887-10	Sequence 10, Appl
44	478	44.3	1366	4	US-09-289-578-10	Sequence 10, Appl
45	477	44.2	504	3	US-09-219-849-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-366-009-6
; Sequence 6, Application US/09366009
; Patent No. 6426042
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6426042uto
; Hashino, Kimikazu
; Kato, Ikunoshin
TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSER: WEISER & ASSOCIATES
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/366,009
FILING DATE: 02-Aug-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,156
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 294382/1995
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: JP 051847/1996
FILING DATE: 08-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 977.6507P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

```
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-366-009-6

Query Match          100.0%; Score 1079; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.6e-76;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKGEDEGDFGFKGDMGIKDRGEIGPPRGEDGPPGKRGCGPNDGPGPL 60
DB 1 GIRGLKGTGKGEDEGDFGFKGDMGIKDRGEIGPPRGEDGPPGKRGCGPNDGPGPL 60
QY 61 GPPGEGKGLGVPLGPGYPRQPKSGISGPPFGGANGKGGRTGPKGPPRGQPGTGP 120
DB 61 GPPGEGKGLGVPLGPGYPRQPKSGISGPPFGGANGKGGRTGPKGPPRGQPGTGP 120
QY 121 GERGPRGITGKPGKNSGGDGAPGPPGERGNGCGTGGPPGKGGPDKGLPGHP 180
DB 121 GERGPRGITGKPGKNSGGDGAPGPPGERGNGCGTGGPPGKGGPDKGLPGHP 180
QY 181 GORGET 186
DB 181 GORGET 186

RESULT 2
US-08-809-156B-6
; Sequence 6, Application US/08809156B
; Patent No. 6472204
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; APPLICANT: Uemori, Takashi
; APPLICANT: Ueno, Takashi
; APPLICANT: Koyama, No. 6472204uto
; APPLICANT: Hashino, Kimikazu
; APPLICANT: Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; TITLE OF INVENTION: CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,156B
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03254
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; LENGTH: 186 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-809-156B-6

Query Match          100.0%; Score 1079; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.6e-76;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKGEDEGDFGFKGDMGIKDRGEIGPPRGEDGPPGKRGCGPNDGPGPL 60
DB 1 GIRGLKGTGKGEDEGDFGFKGDMGIKDRGEIGPPRGEDGPPGKRGCGPNDGPGPL 60
QY 61 GPPGEGKGLGVPLGPGYPRQPKSGISGPPFGGANGKGGRTGPKGPPRGQPGTGP 120
DB 61 GPPGEGKGLGVPLGPGYPRQPKSGISGPPFGGANGKGGRTGPKGPPRGQPGTGP 120
QY 121 GERGPRGITGKPGKNSGGDGAPGPPGERGNGCGTGGPPGKGGPDKGLPGHP 180
DB 121 GERGPRGITGKPGKNSGGDGAPGPPGERGNGCGTGGPPGKGGPDKGLPGHP 180
QY 181 GORGET 186
DB 181 GORGET 186

RESULT 3
US-08-836-854-19
; Sequence 19, Application US/08836854
; Patent No. 5824547
; GENERAL INFORMATION:
; APPLICANT: HASHINO, Kimikazu
; APPLICANT: MATSUSHITA, Hideyuki
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W. Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,854
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP95/02425
; FILING DATE: 29-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 317721/1994
; FILING DATE: 29-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: HASHINO=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```



```
; MOLECULE TYPE: peptide
; US-08-836-854-19
Query Match 100.0%; Score 1079; DB 2; Length 464;
Best Local Similarity 100.0%; Pred. No. 3.8e-76;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKEGEDGPGFKGDMGIKGRGEIGPPGPRGEDGPGPKGRGGNGDPGPL 60
Db 279 GIRGLKGTGKEGEDGPGFKGDMGIKGRGEIGPPGPRGEDGPGPKGRGGNGDPGPL 338
QY 61 GPPGEKGLGVPLGPGVGRQGPKGSGIRPFGFANGKGGRTGPKGPRGGRGTGPR 120
Db 339 GPPGEKGLGVPLGPGVGRQGPKGSGIRPFGFANGKGGRTGPKGPRGGRGTGPR 398
QY 121 GERGPRGITGKPGKNGSGDGPAGPPGERGPNPGQPTGFPKGPBGKDXGLPGHP 180
Db 399 GERGPRGITGKPGKNGSGDGPAGPPGERGPNPGQPTGFPKGPBGKDXGLPGHP 458
QY 181 GORGET 186
Db 459 GORGET 464

RESULT 4
US-09-366-009-7
; Sequence 7, Application US/09366009
; Patent No. 6426042
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6426042uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
```

```
; MOLECULE TYPE: peptide
; US-09-366-009-7
Query Match 100.0%; Score 1079; DB 4; Length 464;
Best Local Similarity 100.0%; Pred. No. 3.8e-76;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKEGEDGPGFKGDMGIKGRGEIGPPGPRGEDGPGPKGRGGNGDPGPL 60
Db 279 GIRGLKGTGKEGEDGPGFKGDMGIKGRGEIGPPGPRGEDGPGPKGRGGNGDPGPL 338
QY 61 GPPGEKGLGVPLGPGVGRQGPKGSGIRPFGFANGKGGRTGPKGPRGGRGTGPR 120
Db 339 GPPGEKGLGVPLGPGVGRQGPKGSGIRPFGFANGKGGRTGPKGPRGGRGTGPR 398
QY 121 GERGPRGITGKPGKNGSGDGPAGPPGERGPNPGQPTGFPKGPBGKDXGLPGHP 180
Db 399 GERGPRGITGKPGKNGSGDGPAGPPGERGPNPGQPTGFPKGPBGKDXGLPGHP 458
QY 181 GORGET 186
Db 459 GORGET 464

RESULT 5
US-08-809-156B-7
; Sequence 7, Application US/08809156B
; Patent No. 6472204
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6472204uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,156B
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03254
; FILING DATE: 07-NOV-1996
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-809-156B-7

Query Match 100.0%; Score 1079; DB 4; Length 464;
Best Local Similarity 100.0%; Pred. No. 3.8e-76;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTKEGEGDFGFKGDMGKIGDRGEIGPPGRGEDGPPGKRGPGNGDPGL 60
Db 279 GIRGLKGTKEGEGDFGFKGDMGKIGDRGEIGPPGRGEDGPPGKRGPGNGDPGL 338
QY 61 GPPGKGLGVPLGYPGRQPKSGIGFPFPGANGKGGRTGKPGRQGRPTGPR 120
Db 339 GPPGKGLGVPLGYPGRQPKSGIGFPFPGANGKGGRTGKPGRQGRPTGPR 398
QY 121 GERGRGITGKPGKNSGGDPAGPPGERGPNPGQPTGTPGPKGPPGPKDGLPGHP 180
Db 399 GERGRGITGKPGKNSGGDPAGPPGERGPNPGQPTGTPGPKGPPGPKDGLPGHP 458
QY 181 GORGET 186
Db 459 GORGET 464

RESULT 6
US-09-366-009-8
; Sequence 8, Application US/09366009
; Patent No. 6426042
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6426042uto
; Hashino, Kimikazu
; Kato, Ikunoehin
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESS: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-366-009-8

Query Match 99.1%; Score 1069; DB 4; Length 489;
Best Local Similarity 100.0%; Pred. No. 2.4e-75;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTKEGEGDFGFKGDMGKIGDRGEIGPPGRGEDGPPGKRGPGNGDPGL 60
Db 279 GIRGLKGTKEGEGDFGFKGDMGKIGDRGEIGPPGRGEDGPPGKRGPGNGDPGL 338
QY 61 GPPGKGLGVPLGYPGRQPKSGIGFPFPGANGKGGRTGKPGRQGRPTGPR 120
Db 339 GPPGKGLGVPLGYPGRQPKSGIGFPFPGANGKGGRTGKPGRQGRPTGPR 398
QY 121 GERGRGITGKPGKNSGGDPAGPPGERGPNPGQPTGTPGPKGPPGPKDGLPGHP 180
Db 399 GERGRGITGKPGKNSGGDPAGPPGERGPNPGQPTGTPGPKGPPGPKDGLPGHP 458
QY 181 GORGET 184
Db 459 GORGET 462

RESULT 7
US-08-809-156B-8
; Sequence 8, Application US/08809156B
; Patent No. 6472204
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; APPLICANT: Uemori, Takashi
; APPLICANT: Ueno, Takashi
; APPLICANT: Koyama, No. 6472204uto
; APPLICANT: Hashino, Kimikazu
; APPLICANT: Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,156B
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03254
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
```


RESULT 14

US-08-642-255-32
; Sequence 32, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A5556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-642-255-32

Query Match 45.2%; Score 488; DB 1; Length 330;
Best Local Similarity 52.1%; Pred. No. 1e-30;
Matches 99; Conservative 1; Mismatches 84; Indels 6; Gaps 1;
QY 1 GIRLKGTKGKGDGDFGFKDMGIKDRGIEGPPRGEDGPGPKRGPGNDPGPL 60
Db 61 GPPGAGPVGSGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 120
QY 61 GPPGKGLGVPLGYPGRQPKSGIGPFPFGANGKGGRTGPKGPRGQRTGPR 120
Db 121 GPPGPPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 180
QY 121 GERGP-----RGITGKPKGNSGDDGAGPPGERGNGPGQPTGFGPKGPPGPKD 174
Db 181 GPAGVGSAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 240
QY 175 GLPGHPGQRG 184
Db 241 GPPGAGPPG 250

RESULT 15

US-07-609-716-65
; Sequence 65, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph

; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-609-716-65

Query Match 45.2%; Score 488; DB 1; Length 408;
Best Local Similarity 52.1%; Pred. No. 1.2e-30;
Matches 99; Conservative 1; Mismatches 84; Indels 6; Gaps 1;
QY 1 GIRLKGTKGKGDGDFGFKDMGIKDRGIEGPPRGEDGPGPKRGPGNDPGPL 60
Db 61 GPPGAGPVGSGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 120
QY 61 GPPGKGLGVPLGYPGRQPKSGIGPFPFGANGKGGRTGPKGPRGQRTGPR 120
Db 121 GPPGPPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 180
QY 121 GERGP-----RGITGKPKGNSGDDGAGPPGERGNGPGQPTGFGPKGPPGPKD 174
Db 181 GPAGVGSAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 240
QY 175 GLPGHPGQRG 184
Db 241 GPPGAGPPG 250
Search completed: May 3, 2004, 13:12:35
Job time : 10.5459 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:11:14 ; Search time 22.2737 Seconds
(without alignments)
2314.713 Million cell updates/sec

Title: US-09-775-964-6
Perfect score: 1079
Sequence: 1 GRGLKGTGKGGDGPFG.....PPGPPKGDLGPHGQGRGET 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1079	100.0	186	10 US-09-775-964-6	Sequence 6, Appli
2	1079	100.0	464	10 US-09-775-964-7	Sequence 7, Appli
3	1069	99.1	489	10 US-09-775-964-8	Sequence 8, Appli
4	930	86.2	1806	9 US-09-919-497-56	Sequence 56, Appli
5	930	86.2	1806	12 US-10-058-270A-122	Sequence 122, App
6	741	68.7	1739	10 US-09-795-061-2	Sequence 2, Appli
7	737	68.3	1745	10 US-09-795-061-4	Sequence 4, Appli
8	509.5	47.2	1466	15 US-10-402-089-4	Sequence 4, Appli
9	509.5	47.2	1466	15 US-10-402-089-6	Sequence 6, Appli
10	509.5	47.2	1466	15 US-10-402-072A-4	Sequence 4, Appli
11	509.5	47.2	1466	15 US-10-402-072A-6	Sequence 6, Appli
12	492.5	45.6	1466	15 US-10-402-089-12	Sequence 12, Appli
13	492.5	45.6	1466	15 US-10-402-072A-12	Sequence 12, Appli
14	491	45.5	1496	14 US-10-177-293-70	Sequence 70, Appli
15	491	45.5	1496	14 US-10-301-822-35	Sequence 35, Appli

16	491	45.5	1496	15	US-10-236-031B-74	Sequence 74, Appli
17	487	45.1	1449	15	US-10-402-089-8	Sequence 8, Appli
18	487	45.1	1449	15	US-10-402-072A-8	Sequence 8, Appli
19	486.5	45.1	1078	14	US-10-058-124-21	Sequence 21, Appli
20	486.5	45.1	1466	12	US-09-918-715-226	Sequence 226, App
21	486.5	45.1	1466	12	US-10-257-021-72	Sequence 72, Appli
22	486.5	45.1	1466	14	US-10-177-293-68	Sequence 68, Appli
23	486.5	45.1	1466	14	US-10-301-822-33	Sequence 33, Appli
24	486	45.0	1366	12	US-10-257-021-76	Sequence 76, Appli
25	486	45.0	1366	14	US-10-058-124-19	Sequence 19, Appli
26	486	45.0	1366	15	US-10-402-089-10	Sequence 10, Appli
27	486	45.0	1366	15	US-10-402-072A-10	Sequence 10, Appli
28	486	45.0	1669	12	US-09-918-715-252	Sequence 252, App
29	486	45.0	1669	15	US-10-372-683-8	Sequence 8, Appli
30	484	44.9	1497	14	US-10-157-031-128	Sequence 128, App
31	483.5	44.8	1212	15	US-10-331-496A-26	Sequence 26, Appli
32	480	44.5	1626	14	US-10-202-167-2	Sequence 2, Appli
33	479.5	44.4	638	15	US-10-138-588-42	Sequence 42, Appli
34	479.5	44.4	1516	14	US-10-060-036-166	Sequence 166, App
35	479.5	44.4	1516	15	US-10-431-642-3	Sequence 3, Appli
36	478.5	44.3	510	12	US-10-232-175-26	Sequence 26, Appli
37	478.5	44.3	638	13	US-10-001-887-108	Sequence 108, App
38	478.5	44.3	662	12	US-10-232-175-33	Sequence 33, Appli
39	478.5	44.3	703	14	US-10-219-449-4	Sequence 4, Appli
40	478.5	44.3	714	14	US-10-233-885-44	Sequence 44, Appli
41	478.5	44.3	714	14	US-10-231-581-44	Sequence 44, Appli
42	478.5	44.3	714	15	US-10-326-508A-15	Sequence 15, Appli
43	478.5	44.3	717	14	US-10-219-449-2	Sequence 2, Appli
44	478.5	44.3	733	15	US-10-138-588-40	Sequence 40, Appli
45	478.5	44.3	1341	14	US-10-058-124-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1

US-09-775-964-6
; Sequence 6, Application US/09775964
; Publication NO. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. US20030087437A1uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSER: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/775,964
FILING DATE: 20-Feb-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/366,009
FILING DATE: 02-Aug-1999
APPLICATION NUMBER: 08/809,156
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 294382/1995
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: JP 051847/1996


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-775-964-8

Query Match      99.1%; Score 1069; DB 10; Length 489;
Best Local Similarity 100.0%; Pred. No. 2.8e-64;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPPGPRGEDGEGPKGSGGNGDGPPL 60
DB      279 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPPGPRGEDGEGPKGSGGNGDGPPL 338

QY      61 GPPGEKGLGVPLGPGYGRQPKSGISGFFPGFPGANGKGGRTGPKGPRGGRGPTGPR 120
DB      339 GPPGEKGLGVPLGPGYGRQPKSGISGFFPGFPGANGKGGRTGPKGPRGGRGPTGPR 398

QY      121 GERPRGITGKPKGKNSGGDGPAGPPGERGPNQPGTFFPKGPPGPKGDKGLPGHP 180
DB      399 GERPRGITGKPKGKNSGGDGPAGPPGERGPNQPGTFFPKGPPGPKGDKGLPGHP 458

QY      181 GQRG 184
DB      459 GQRG 462

RESULT 4
US-09-919-497-56
; Sequence 56, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (758)..(758)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: UNSURE
; LOCATION: (809)..(809)
; OTHER INFORMATION: Xaa = any amino acid
US-09-919-497-56

Query Match      99.1%; Score 1069; DB 10; Length 489;
Best Local Similarity 100.0%; Pred. No. 2.8e-64;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPPGPRGEDGEGPKGSGGNGDGPPL 60
DB      279 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPPGPRGEDGEGPKGSGGNGDGPPL 338

QY      61 GPPGEKGLGVPLGPGYGRQPKSGISGFFPGFPGANGKGGRTGPKGPRGGRGPTGPR 120
DB      339 GPPGEKGLGVPLGPGYGRQPKSGISGFFPGFPGANGKGGRTGPKGPRGGRGPTGPR 398

QY      121 GERPRGITGKPKGKNSGGDGPAGPPGERGPNQPGTFFPKGPPGPKGDKGLPGHP 180
DB      399 GERPRGITGKPKGKNSGGDGPAGPPGERGPNQPGTFFPKGPPGPKGDKGLPGHP 458

QY      181 GQRG 184
DB      459 GQRG 462

RESULT 5
US-10-058-270A-122
; Sequence 122, Application US/10058270A
; Publication No. US20040029114A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Afar, Daniel
; APPLICANT: Eos biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and
; FILE REFERENCE: 018501-005210US
; CURRENT APPLICATION NUMBER: US/10/058,270A
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: US 60/263,965
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: US 60/265,928
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 09/829,472
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/282,698
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/288,590
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 122
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1806)
; OTHER INFORMATION: Xaa = any amino acid
US-10-058-270A-122

Query Match      86.2%; Score 930; DB 12; Length 1806;
Best Local Similarity 86.0%; Pred. No. 1.8e-54;
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

QY      1 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPPGPRGEDGEGPKGSGGNGDGPPL 60
DB      769 GVRGLKSGKSGEDGPGFKGDMGLKDRGEVQIGPRGXDGPEGPKGRAGTGDGPFS 828

QY      61 GPPGEKGLGVPLGPGYGRQPKSGISGFFPGFPGANGKGGRTGPKGPRGGRGPTGPR 120
DB      829 GQAGEKGLGVPLGPGYGRQPKSGISGFFPGFPGANGKGGRTGPKGPRGGRGPTGPR 888

QY      121 GERPRGITGKPKGKNSGGDGPAGPPGERGPNQPGTFFPKGPPGPKGDKGLPGHP 180
DB      889 GSRGARGPTGPKGKTSGDGGPPGPPGERGPNQPGTFFPKGPPGPKGDKGLPGHP 948

QY      181 GQRGET 186
DB      949 GQRGET 954

Query Match      86.2%; Score 930; DB 9; Length 1806;
Best Local Similarity 86.0%; Pred. No. 1.8e-54;
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

QY      1 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPPGPRGEDGEGPKGSGGNGDGPPL 60
DB      769 GVRGLKSGKSGEDGPGFKGDMGLKDRGEVQIGPRGXDGPEGPKGRAGTGDGPFS 828

QY      61 GPPGEKGLGVPLGPGYGRQPKSGISGFFPGFPGANGKGGRTGPKGPRGGRGPTGPR 120
DB      829 GQAGEKGLGVPLGPGYGRQPKSGISGFFPGFPGANGKGGRTGPKGPRGGRGPTGPR 888

QY      121 GERPRGITGKPKGKNSGGDGPAGPPGERGPNQPGTFFPKGPPGPKGDKGLPGHP 180
DB      889 GSRGARGPTGPKGKTSGDGGPPGPPGERGPNQPGTFFPKGPPGPKGDKGLPGHP 948

QY      181 GQRGET 186
DB      949 GQRGET 954
```

```
Db      889 GSRGARGPTGKCPKTSKSGDGPFGPPGGRGPGQGPVGFPGPKGPPGPPGMRGCPGHP 948
QY      181 GORGET 186
Db      949 GORGET 954

RESULT 6
US-09-795-061-2
; Sequence 2, Application US/09795061
; Publication No. US20030166842A1
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; APPLICANT: Imamura, Yasutada
; TITLE OF INVENTION: Pro-Alpha 3(V) Collagen Genes
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-795-061-2

Query Match      68.7%; Score 741; DB 10; Length 1739;
Best Local Similarity 73.0%; Pred. No. 8.1e-42;
Matches 135; Conservative 10; Mismatches 40; Indels 0; Gaps 0;

QY      1 GIRGLKGTGKEGDEGFGPGFKGDMGIKGRGEIGPPRGEDGPEGPKRGGPNGDGGL 60
Db      718 GNRGLGEGKGERGDEGFGPGFKGDMGIKGRGEIGPPRGEDGPEGPKRGGPNGDGGL 777

QY      61 GPPGEGKGLGVPLGPGYPCGRQPKSGISGFPFGANGKGGRTGPKGPRGQGTGPR 120
Db      778 GAAGEKGLGVPLGPGYPCGRQPKSGISGFPFGANGKGGRTGPKGPRGQGTGPR 837

QY      121 GERGPRGITKPKKNSGGDGPAGPPGGRGNGPGPTGFGPKGPPGPKDGLPGHP 180
Db      838 GDRGQPGATGQPGFKGDMGVGNSGPPGPKGULPGLQGGPPGPGPKGPPGQKRGDISGHP 897

QY      181 GORGE 185
Db      898 GORGE 902

RESULT 7
US-09-795-061-4
; Sequence 4, Application US/09795061
; Publication No. US20030166842A1
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; APPLICANT: Imamura, Yasutada
; TITLE OF INVENTION: Pro-Alpha 3(V) Collagen Genes
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-061-4

Query Match      68.3%; Score 737; DB 10; Length 1745;
Best Local Similarity 72.4%; Pred. No. 1.5e-41;
Matches 134; Conservative 12; Mismatches 39; Indels 0; Gaps 0;

QY      1 GIRGLKGTGKEGDEGFGPGFKGDMGIKGRGETGPPGPRGEDGPEGPKRGGPNGDGGL 60
```

```
Db      719 GNRGLQGEKGEKGBDGPFGFKGDMGVGLKGDQCKPGAPGPRGEDGPEGPKGQAGQAGEGPP 778
QY      61 GPPGEGKGLGVPLGPGYPCGRQPKSGISGFPFGANGKGGRTGPKGPRGQGTGPR 120
Db      779 GSAGEKGLGVPLGPGYPCGRQPKSGISGFPFGPLGPIGEKKSCKTQPGLEGERGPPGSR 838
QY      121 GERGPRGITKPKKNSGGDGPAGPPGGRGNGPGPTGFGPKGPPGPKDGLPGHP 180
Db      839 GERGQPGATGQPGFKGDMGVGDGAPGIFGEKGLPGLQGGPPGPPGPKGPPGHQKDGREFHP 898

QY      181 GORGE 185
Db      899 GORGE 903

RESULT 8
US-10-402-089-4
; Sequence 4, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Bos Taurus
US-10-402-089-4

Query Match      47.2%; Score 509.5; DB 15; Length 1466;
Best Local Similarity 38.5%; Pred. No. 2.3e-26;
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;

QY      1 GIRGLKGT-----KGEKGEDGFPFKGDMGIKGRGEI---GPPGPRGED 42
Db      639 GLQGLPGTSGPPGNGKFGEPFKGEAGAPGIPGCKGSDGAPGERGPPGAGGPPGPGGA 698
QY      43 GPEGPKG-----RGGPN----- 54
Db      699 GPPGEGKGAAGPPPGPSAGTPOGMPGERGGGPGPKGDKGEGSSGVDCAPGKD 758

QY      55 -----GDPGLGPPGEGKGLGVPLGYP-----GRQPKSGISGFPFGPGAN 96
Db      759 GPRGPTGIPGPPGAGQFGDKGESGAPGVPIAGPRGGGGERGEGQPGPAGFPAGPQN 818

QY      97 GEKGRGTPGKPPRGQ-----RGTPRGERGPRGIGTKPKPKNSGSD 141
Db      819 GEPGAKGERGAPGEKGEKGGPPCAAGPAGGSGPAGPGPGQVKGERGSGPGPAAGFPQGR 878

QY      142 GPAGPPGGRGNGPGPTGFGPKGPPGPPGKDGLPHPGQRG 184
Db      879 GPPGPPGSGNPNPPPGSSGAPFKDGGPPGPNAGPAGSGFISG 921

RESULT 9
US-10-402-089-6
; Sequence 6, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
```

FILE REFERENCE: FP0402.3 CON
CURRENT APPLICATION NUMBER: US/10/402,089
CURRENT FILING DATE: 2003-03-26
PRIOR APPLICATION NUMBER: US 09/709,700
PRIOR FILING DATE: 2000-11-10
NUMBER OF SEQ ID NOS: 72
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 1466
TYPE: PRT
ORGANISM: Sus scrofa
US-10-402-089-6

Query Match 47.2%; Score 509.5; DB 15; Length 1466;
Best Local Similarity 38.5%; Pred. No. 2.3e-26;
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;
1 GIRLKGKT-----KGEKGDGFGPKGDMGKIGDRGEI---GPPGPRGED 42
639 GLOGLPTGPPGNGKPGKGEAGAPGIPGKGDGSGAPGERGPPGAGGPPGPRGGA 698
43 GPEGPKG-----RGGPN-----54
699 GPPGEGKGAAGPPGPPGSAFTPLQGMPPGPGPKGDKGEPGSSGVDGAPGKD 758
55 -----GDPGLGPPGKGLGVPLPGYP-----GRQPKGSIGPPGPPGAN 96
759 GPRGPTGPIGPPGAGPGDKGSGAPGVPIAGPRGGPGERGQQGPPGAGPPGAPGQN 818
97 GEKGGRTGPKGPPRQ-----RGPTGPRGERGPRGITGKPKGKNSGGD 141
819 GEPKAGGERGAPGKGEKGGPPGAAGGAGGPPGPPGVKGERSGPGGAGGPPGGR 878
142 GPAGPPGERGPPGQGTGTPGPKGPPGPKDGLFGHPGQRG 184
879 GPPGPPGSGNPPGPPGSSGAPGKDGPPGPPGSGNAPGSPGISG 921

RESULT 10
US-10-402-072A-4
Sequence 4, Application US/10402072A
Publication No. US20040018592A1
GENERAL INFORMATION:
APPLICANT: Bell, Marcum P.
APPLICANT: Neff, Thomas B.
APPLICANT: Polarek, James W.
APPLICANT: Seeley, Todd W.
TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
FILE REFERENCE: FP0402.2 CON
CURRENT APPLICATION NUMBER: US/10/402,072A
CURRENT FILING DATE: 2003-03-26
PRIOR APPLICATION NUMBER: US 09/709,700
PRIOR FILING DATE: 2000-11-10
NUMBER OF SEQ ID NOS: 72
SOFTWARE: PatentIn version 3.2
SEQ ID NO 4
LENGTH: 1466
TYPE: PRT
ORGANISM: Bos Taurus
US-10-402-072A-4

Query Match 47.2%; Score 509.5; DB 15; Length 1466;
Best Local Similarity 38.5%; Pred. No. 2.3e-26;
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;
1 GIRLKGKT-----KGEKGDGFGPKGDMGKIGDRGEI---GPPGPRGED 42
639 GLOGLPTGPPGNGKPGKGEAGAPGIPGKGDGSGAPGERGPPGAGGPPGPRGGA 698
43 GPEGPKG-----RGGPN-----54
699 GPPGEGKGAAGPPGPPGSAFTPLQGMPPGPGPKGDKGEPGSSGVDGAPGKD 758

55 -----GDPGLGPPGKGLGVPLPGYP-----GRQPKGSIGPPGPPGAN 96
759 GPRGPTGPIGPPGAGPGDKGSGAPGVPIAGPRGGPGERGQQGPPGAGPPGAPGQN 818
97 GEKGGRTGPKGPPRQ-----RGPTGPRGERGPRGITGKPKGKNSGGD 141
819 GEPKAGGERGAPGKGEKGGPPGAAGGAGGPPGPPGVKGERSGPGGAGGPPGGR 878
142 GPAGPPGERGPPGQGTGTPGPKGPPGPKDGLFGHPGQRG 184
879 GPPGPPGSGNPPGPPGSSGAPGKDGPPGPPGSGNAPGSPGISG 921

RESULT 11
US-10-402-072A-6
Sequence 6, Application US/10402072A
Publication No. US20040018592A1
GENERAL INFORMATION:
APPLICANT: Bell, Marcum P.
APPLICANT: Neff, Thomas B.
APPLICANT: Polarek, James W.
APPLICANT: Seeley, Todd W.
TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
FILE REFERENCE: FP0402.2 CON
CURRENT APPLICATION NUMBER: US/10/402,072A
CURRENT FILING DATE: 2003-03-26
PRIOR APPLICATION NUMBER: US 09/709,700
PRIOR FILING DATE: 2000-11-10
NUMBER OF SEQ ID NOS: 72
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 1466
TYPE: PRT
ORGANISM: Sus scrofa
US-10-402-072A-6

Query Match 47.2%; Score 509.5; DB 15; Length 1466;
Best Local Similarity 38.5%; Pred. No. 2.3e-26;
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;
1 GIRLKGKT-----KGEKGDGFGPKGDMGKIGDRGEI---GPPGPRGED 42
639 GLOGLPTGPPGNGKPGKGEAGAPGIPGKGDGSGAPGERGPPGAGGPPGPRGGA 698
43 GPEGPKG-----RGGPN-----54
699 GPPGEGKGAAGPPGPPGSAFTPLQGMPPGPGPKGDKGEPGSSGVDGAPGKD 758
55 -----GDPGLGPPGKGLGVPLPGYP-----GRQPKGSIGPPGPPGAN 96
759 GPRGPTGPIGPPGAGPGDKGSGAPGVPIAGPRGGPGERGQQGPPGAGPPGAPGQN 818
97 GEKGGRTGPKGPPRQ-----RGPTGPRGERGPRGITGKPKGKNSGGD 141
819 GEPKAGGERGAPGKGEKGGPPGAAGGAGGPPGPPGVKGERSGPGGAGGPPGGR 878
142 GPAGPPGERGPPGQGTGTPGPKGPPGPKDGLFGHPGQRG 184
879 GPPGPPGSGNPPGPPGSSGAPGKDGPPGPPGSGNAPGSPGISG 921

RESULT 12
US-10-402-089-12
Sequence 12, Application US/10402089
Publication No. US20040005663A1
GENERAL INFORMATION:
APPLICANT: Bell, Marcum P.
APPLICANT: Neff, Thomas B.
APPLICANT: Polarek, James W.
APPLICANT: Seeley, Todd W.
TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
FILE REFERENCE: FP0402.3 CON
CURRENT APPLICATION NUMBER: US/10/402,089

; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PR
; ORGANISM: Sus scrofa
US-10-402-089-12

Query Match 45.6%; Score 492.5; DB 15; Length 1466;
Best Local Similarity 41.9%; Pred. No. 3.2e-25;
Matches 104; Conservative 11; Mismatches 70; Indels 63; Gaps 5;

QY 1 GIRGLKGTGKGEDEGPFPGKDMGKIDRGEIGPPRGEDGPEGPKRGGPNGDGPGPL 60
DB 574 GQGVNGVFPFGKNDGAPKNGRGGPGGFLPGPPKNGETGPGQPPGPTGPGDGKGT 633
QY 61 GPPGEKGLGVPLP-----GVPGRQGFKSGIGFPFGANGEXKGRGTGPKP----- 108
DB 634 GPPGQGLQGLPFTSGPPGKNGKPEGPKGEAGAGIIPGKGDGSGAPGERPPGAVGPS 693
QY 109 GPRGQRGPTGPRGERP-----RGITKPKPKGNS-----GGDGPA 144
DB 694 GPRGGAGPFGPEGKPGAGPPGPPGAAAGTFLQGMPPGERGSGGPGPKDKDGPGGSGAD 753
QY 145 GPPGERGPNQPGTGFPGPKGPP-----GPPKDGLP 177
DB 754 GAPGKDGPRGPTGPIGPPGPAQPGQDKESGAPFLGIAGPRGGGERGHEGPPGAGFP 813
QY 178 GHPQORGE 185
DB 814 GAPQNGE 821

RESULT 13
US-10-402-072A-12
; Sequence 12, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PR
; ORGANISM: Sus scrofa
US-10-402-072A-12

Query Match 45.6%; Score 492.5; DB 15; Length 1466;
Best Local Similarity 41.9%; Pred. No. 3.2e-25;
Matches 104; Conservative 11; Mismatches 70; Indels 63; Gaps 5;

QY 1 GIRGLKGTGKGEDEGPFPGKDMGKIDRGEIGPPRGEDGPEGPKRGGPNGDGPGPL 60
DB 574 GQGVNGVFPFGKNDGAPKNGRGGPGGFLPGPPKNGETGPGQPPGPTGPGDGKGT 633
QY 61 GPPGEKGLGVPLP-----GVPGRQGFKSGIGFPFGANGEXKGRGTGPKP----- 108
DB 634 GPPGQGLQGLPFTSGPPGKNGKPEGPKGEAGAGIIPGKGDGSGAPGERPPGAVGPS 693
QY 109 GPRGQRGPTGPRGERP-----RGITKPKPKGNS-----GGDGPA 144

Db 694 GPRGAGPPGPEGKGAGPPPPGAAAGTFLQMPGCGSGGPGPKDKDGPGGSGAD 753
QY 145 GPPGERGPNQPGTGFPGPKGPP-----GPPKDGLP 177
DB 754 GAPGKDGPRGPTGPIGPPGPAQPGQDKESGAPFLGIAGPRGGGERGHEGPPGAGFP 813
QY 178 GHPQORGE 185
DB 814 GAPQNGE 821

RESULT 14
US-10-177-293-70
; Sequence 70, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1496
; TYPE: PR
; ORGANISM: Homo sapiens
US-10-177-293-70

Query Match 45.5%; Score 491; DB 14; Length 1496;
Best Local Similarity 51.4%; Pred. No. 4.1e-25;
Matches 95; Conservative 12; Mismatches 78; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKGEDEGPFPGKDMGKIDRGEIGPPRGEDGPEGPKRGGPNGDGPGPL 60
DB 291 GAGPLGLKGRHKGLESGKGEVGA PGSGKGEAGTGP MGAMGLGPRGMFGERGLGFPQ 350
QY 61 GPPGEKGLGVPLPDPYFGROGPKSGI GPPFGPANGEXKGRGTGPKGPRGQGTGPR 120
DB 351 GAPGQRGAHMPKDFGPMGLGIPGSSGFP GPNMGKGEAGTGA RGPPEGQGTGPP 410
QY 121 GERGPRGITGKPGPKGNSGGDPAGPGERPN GQQGTGPPGKPGPPGPKDGLPGHP 180

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:06:19 ; Search time 8.09953 Seconds
(without alignments)
2208.970 Million cell updates/sec

Title: US-09-775-964-6
Perfect score: 1079
Sequence: 1 GIRGLKGTGKEKCEDGPPGF.....PPGPPKDGLPQHPGQGET 186

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1079	100.0	1838	1 CGHUIV	collagen alpha 1(V)
2	1071	99.3	1843	2 S18803	collagen alpha 1(V)
3	953	88.3	1024	2 S18251	collagen alpha 1(X)
4	930	86.2	1806	1 CGHUIE	collagen alpha 1(X)
5	873	80.9	1546	1 CGH02E	collagen alpha 2(X)
6	509.5	47.2	1049	1 CGB07S	collagen alpha 1(I)
7	501	46.4	1669	1 CGM34B	collagen alpha 1(I)
8	498.5	46.2	1691	1 S22917	collagen alpha 1(I)
9	497.5	46.1	888	2 S22971	collagen alpha 1(X)
10	494	45.8	674	2 S23297	collagen alpha 1(X)
11	491	45.5	1496	1 CGH02V	collagen alpha 2(V)
12	489.5	45.4	1464	2 S59856	collagen alpha 1(I)
13	488.5	45.3	1019	1 A32856	collagen alpha 1(V)
14	488.5	45.3	1492	2 A40333	collagen alpha 1(I)
15	486.5	45.1	1466	1 CGH07L	collagen alpha 1(I)
16	486	45.0	1669	1 CGH04B	collagen alpha 1(I)
17	485.5	45.0	674	2 S13301	collagen alpha 1(X)
18	484.5	44.9	1486	1 B40333	collagen alpha 1(I)
19	484	44.9	1532	2 A61262	collagen alpha 1(X)
20	483	44.8	964	1 CGCH2S	collagen alpha 2(I)
21	481.5	44.6	1146	2 A38587	collagen, cornea-s
22	480.5	44.5	920	2 A45748	collagen alpha 1(V)
23	479.5	44.4	471	2 A39024	collagen alpha 3(I)
24	479.5	44.4	671	1 CGRT1S	collagen alpha 1(I)
25	479.5	44.4	886	2 I50694	collagen alpha 1(I)
26	479.5	44.4	1418	2 T45467	collagen alpha 1(I)
27	479	44.4	402	1 CGB02S	collagen alpha 2(I)
28	479	44.4	488	2 A27353	collagen alpha 1(I)
29	478.5	44.3	1464	1 CGHUI5	collagen alpha 1(I)

RESULT 1

CGHUIV

collagen alpha 1(V) chain precursor - human

N;Alternate names: procollagen alpha 1(V) chain

C;Species: Homo sapiens (man)

C;Date: 22-Nov-1993 #sequence revision 03-Oct-1995 #text change 16-Jun-2000

C;Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665

R;GreenSpan, D.S.; Cheng, W.; Hoffman, G.G.

J. Biol. Chem. 266, 24727-24733, 1991

A;Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of

A;Reference number: S18802; MUID:92105142; PMID:1722213

A;Accession: S18802

A;Molecule type: mRNA

A;Residues: 1-1838 <GRE>

A;Cross-references: GB:M76729; NID:gl89519; PIDN:AAAS9993.1; PID:gl89520

R;Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaoi, Y.; Kato, I.

J. Biol. Chem. 266, 13124-13129, 1991

A;Title: Complete primary structure of human collagen alpha-1(V) chain.

A;Reference number: S16024; MUID:91302336; PMID:2071595

A;Accession: S16024

A;Molecule type: mRNA

A;Residues: 1-81, 'QU', '84-389, 'A', '391-676, 'K', '678-1294, 'RS', '1297, 'RS', '1300-1553, 'R', '1555

A;Cross-references: GB:D90279; NID:G219509; PIDN:BAAI4323.1; PID:G219510

A;Note: parts of this sequence were determined by protein sequencing

R;Yaoi, Y.; Hashimoto, K.; Takahara, K.; Kato, I.

Exp. Cell Res. 194, 180-185, 1991

A;Title: Insulin binds to type V collagen with retention of mitogenic activity.

A;Reference number: A61142; MUID:91224163; PMID:1709100

A;Accession: A61142

A;Molecule type: protein

A;Residues: 823-824, 'X', '836-842 <YAO>

A;Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyanogen

R;Yaoi, Y.; Hashimoto, K.; Koitabashi, H.; Takahara, K.; Ito, M.; Kato, I.

Biochim. Biophys. Acta 1035, 139-145, 1990

A;Title: Primary structure of the heparin-binding site of type V collagen.

A;Reference number: S11303; MUID:90366601; PMID:2203476

A;Accession: S11303

A;Molecule type: protein

A;Residues: 823-824, 'X', '826-848, 'I', '850-851, 'P', '853, 'PR', '856-893, 'D', '895-932, 'X', '934-951

A;Note: the residues designated 'X' are probably glycosylated hydroxylysine; this sequen

R;Seyer, J.M.; Kang, A.H.

Arch. Biochem. Biophys. 271, 120-129, 1989

A;Title: Covalent structure of collagen: amino acid sequence of three cyanogen bromide-

A;Reference number: S03978; MUID:89227189; PMID:2496661

A;Accession: S03978

A;Molecule type: protein

A;Residues: 621-640, 'G', '642-649, 'L', '651-662, 'E', '664-667, 'Q', '669-676, 'Q', '678-683, 'P', '685

<SE>

A;Note: there are a number of inconsistencies between the sequences in figures 6 and 7;

R;Moradi-Ameli, M.; Rousseau, J.C.; Klemen, J.P.; Champliand, M.F.; Boutillon, M.M.; Be

Eur. J. Biochem. 221, 987-995, 1994

A;Title: Diversity in the processing events at the N-terminus of type-V collagen.

A;Reference number: S43642; MUID:94237164; PMID:8181482
A;Accession: S43642
A;Molecule type: protein
A;Residues: 565-576;756-758,'X',760-763,'X',765-772;1012-1029;1219-1232;1465-1474,'X',14
R;Fessler, L.I.; Brosh, S.; Chapin, S.; Fessler, J.H.
J. Biol. Chem. 261, 5034-5040, 1986
A;Title: Tyrosine sulfation in precursors of collagen V.
A;Reference number: A56977; MUID:86168226; PMID:3082875
A;Contents: annotation; identification of tyrosine sulfate in the amino-terminal propept
R;Lee, S.; Greenspan, D.S.
Biochem. J. 310, 15-22, 1995
A;Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).
A;Reference number: S58665; MUID:95374437; PMID:7646438
A;Accession: S58665
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-36 <LE>
A;Cross-references: GB:I38808; NID:gl020325; PIDN:AA79853.1; PID:gl020326
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
are 5-hydroxylated and subsequently O-glycosylated.
C;Comment: A long form of the mature protein containing part of the amino-terminal prope
ile the heterotrimers are probably processed to the long form.
C;Genetics:
A;Gene: GDB:COL5A1
A;Cross-references: GDB:I31457; OMIM:120215
A;Map position: 9q34.2-9q34.3
C;Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer of t
alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
engh, is formed with desmosine cross-links made from lysine and allysine residues
C;Function:
A;Description: structural component of extracellular fibrous polymer associated with cel
A;Note: may play a role in controlling the lateral growth of collagen I fibrils
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
F;1-37/Domain: signal sequence #status predicted <SIG>
F;36-261/Domain: PARP-like #status predicted <PAR>
F;38-541/Domain: amino-terminal propeptide #status predicted <PRO>
F;542-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>
F;542-558/Region: amino-terminal nonhelical telopeptide
F;559-1572/Region: helical
F;545-647/Region: cell attachment (R-G-D) motif
F;663-665/Region: cell attachment (R-G-D) motif
F;897-929/Region: heparin binding
F;1573-1605/Region: carboxyl-terminal nonhelical telopeptide
F;1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;62-244,183-237/Disulfide bonds: #status predicted
F;159,176,385,1672,1741/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;234,236,240,262,263,273,274,275,277,279,280,338,340,346,347,352,357,416,417,420,421/Bi
F;535/Modified site: allysine (Lys) #status predicted
F;541-542/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F;542/Modified site: Pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
F;570,576,621,639,648,654,657,675,678,690,693,696,705,717,720,726,732,741,750,753,756,78
site: 4-hydroxyproline (Pro) #status experimental
F;627,642,687,708,744,774,795,804,807,810,819,825,846,864,882,897/Modified site: 5-hydro
F;627,642,687,774,795,804,807,810,819,825,846,864,882,897,1482/Binding site: carbohydrat
F;708,744/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;1482/Modified site: 5-hydroxylysine (Lys) #status predicted
F;1605-1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted
F;1639,1645,1662,1671/Disulfide bonds: interchain #status predicted
F;1680-1835,1746-1789/Disulfide bonds: #status predicted

Query Match 100.0%; Score 1079; DB 1; Length 1838;
Best Local Similarity 100.0%; Pred. No. 8.2e-60;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKEGDFPGFKGDMGIKGRGEIGPPGPGRGEDGPEGPKRGCGPNDGPGL 60
|||||
DB 799 GIRGLKGTGKEGDFPGFKGDMGIKGRGEIGPPGPGRGEDGPEGPKRGCGPNDGPGL 858
|||||

QY 61 GPPGEKGLGVPLPGYPRGQPKSGIGFPGPGANGKGGRTGPKPGPRGQGTGPR 120
|||||

DB 859 GPPGEKGLGVPLPGYPRGQPKSGIGFPGPGANGKGGRTGPKPGPRGQGTGPR 918
QY 121 GERGPRGITGPKPGKNSGGDGAPGPPGERGPNPGQGTGTPGPKPGPPGPKGKGLPGHP 180
|||||
DB 919 GERGPRGITGPKPGKNSGGDGAPGPPGERGPNPGQGTGTPGPKPGPPGPKGKGLPGHP 978
QY 181 GORGET 186
|||||
DB 979 GORGET 984
|||||

RESULT 2
S18803
collagen alpha 1(V) chain - hamster
C;Species: Cricetinae gen. sp. (hamster)
C;Date: 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change 16-Dec-1998
C;Accession: S18803
R;Greenspan, D.S.; Cheng, W.; Hoffman, G.G.
J. Biol. Chem. 266, 24727-24733, 1991
A;Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of
A;Reference number: S18802; MUID:92105142; PMID:1722213
A;Accession: S18803
A;Status: Preliminary
A;Molecule type: mRNA
A;Residues: 1-1843 <GRE>
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
F;1620-1842/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 99.3%; Score 1071; DB 2; Length 1843;
Best Local Similarity 99.5%; Pred. No. 2.6e-59;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKEGDFPGFKGDMGIKGRGEIGPPGPGRGEDGPEGPKRGCGPNDGPGL 60
804 GIRGLKGTGKEGDFPGFKGDMGIKGRGEIGPPGPGRGEDGPEGPKRGCGPNDGPGL 863
DB 61 GPPGEKGLGVPLPGYPRGQPKSGIGFPGPGANGKGGRTGPKPGPRGQGTGPR 120
864 GPTGEKGLGVPLPGYPRGQPKSGIGFPGPGANGKGGRTGPKPGPRGQGTGPR 923
QY 121 GERGPRGITGPKPGKNSGGDGAPGPPGERGPNPGQGTGTPGPKPGPPGPKGKGLPGHP 180
DB 924 GERGPRGITGPKPGKNSGGDGAPGPPGERGPNPGQGTGTPGPKPGPPGPKGKGLPGHP 983
QY 181 GORGET 186
DB 984 GORGET 989
|||||

RESULT 3
S18251
collagen alpha 1(XI) chain - bovine (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 22-Nov-1993 #sequence_revision 23-Mar-1995 #text_change 02-Jul-1998
A;Accession: S18251; C46662; A56978; S65864; D46662; E46662; G46662; H46662; I46
R;Brown, K.E.; Lawrence, R.; Sonenshein, G.E.
J. Biol. Chem. 266, 23268-23273, 1991
A;Title: Concerted modulation of alpha-1(XI) and alpha-2(V) collagen mRNAs in bovine vas
A;Reference number: S18251; MUID:92078200; PMID:1744123
A;Accession: S18251
A;Molecule type: mRNA
A;Residues: 1-911 <BRO>
R;Mayne, R.; Brewton, R.G.; Mayne, P.M.; Baker, J.R.
J. Biol. Chem. 268, 9381-9386, 1993
A;Title: Isolation and characterization of the chains of type V/type XI collagen present
A;Reference number: A46662; MUID:93252802; PMID:8486632
A;Accession: C46662
A;Molecule type: protein
A;Residues: 347-354;356-363;586-600;912-924;925-961;962-998;999-1024 <MAY>
A;Experimental source: vitreous humor
A;Note: sequence modified after extraction from NCBI backbone
J.Wu, J.J.; Eyre, D.R.
J. Biol. Chem. 270, 18865-18870, 1995

A;Title: Structural analysis of cross-linking domains in cartilage type XI collagen. Ins
A;Reference number: A56978; MUID:95370194; PMID:7642541
A;Accession: A56978
A;Status: preliminary
A;Molecule type: protein
A;Residues: 206-229 <WUA>
R;Niyibizi, C.; Eyre, D.R.
Eur. J. Biochem. 224, 943-950, 1994
A;Title: Structural characteristics of cross-linking sites in type V collagen of bone. C
A;Reference number: S48210; MUID:95010086; PMID:7925418
A;Accession: S65864
A;Molecule type: protein
A;Residues: 'X', 273-298 <NIY>
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: hydroxyproline
F;211,223/Modified site: 4-hydroxyproline (Pro) #status experimental

Query Match 88.3%; Score 953; DB 2; Length 1024;
Best Local Similarity 88.2%; Pred. No. 3.2e-52;
Matches 164; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

Qy 1 GIRLKGKTKGKEDGPPGPKGDMGIRKDRGEIGPPGPRGEDGPEGPKGRGPNNDPGL 60
Db 536 GVRGLKSGKGEKEDGPPGPKGDMGLKGDREVGQVGRGEDGPEGPKGRAGPTGDPGP 595

Qy 61 GPPEKGLGVPLGPGVGRQPKSGTGFPGFPANKEKGRGTPGPKGPRGQGTGPR 120
Db 596 GQAGEKGLGVPLGPGVGRQPKSGTGFPGFPANKEKGRGTPGPKGPRGQGTGPR 655

Qy 121 GERPRGITGKPGKNSGGDGPAGPGRGPNQPGTGFPGPKGPPGPKGDXGLPGHP 180
Db 656 GSRGARGPTGKPGKGTSGGDPGPPGERGPPQGPVGFPGPKGPPGPKGDXGLPGHP 715

Qy 181 GQRGET 186
Db 716 GQRGET 721

RESULT 4
CGHUZE
collagen alpha 1(XI) chain precursor - human
N;Alternate names: procollagen alpha 1(XI) chain
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1990 #sequence_revision 03-Oct-1995 #text_change 08-May-1998
C;Accession: A35239; A31795
R;Yoshioka, H.; Ramirez, F.
J. Biol. Chem. 265, 6423-6426, 1990
A;Title: Pro-alpha1(XI) collagen. Structure of the amino-terminal propeptide and expres
A;Reference number: A35239; MUID:90202924; PMID:1690726
A;Accession: A35239
A;Molecule type: mRNA
A;Residues: 1-558 <YOS>
R;Bernard, M.; Yoshioka, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.;
J. Biol. Chem. 263, 17159-17166, 1988
A;Title: Cloning and sequencing of pro-alpha1(XI) collagen cDNA demonstrates that type X
cartilageous tissue.
A;Reference number: A92689; MUID:89034222; PMID:3182841
A;Accession: A31795
A;Molecule type: DNA; mRNA
A;Residues: 538-1806 <BER>
A;Cross-references: GB:J04177
A;Note: parts of this sequence were determined by protein sequencing
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL11A1; COL16
A;Cross-references: GDB:I20595; OMIM:120280
A;Map position: lp21-1p21
A;Introns: 561/3; 579/3; 597/3; 615/3; 633/3; 648/3; 666/3; 681/3
A;Note: the list of introns is incomplete
C;Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains and one alpha
3(XI) chain (see PIR:CGHU6C), initially linked by disulfide bonds among their carboxyl

xmed with desmosome cross-links made from lysine and allysine residues
C;Function:
A;Description: structural component of extracellular fibrous polymer associated with ce
A;Note: may play a role in controlling the lateral growth of collagen II fibrils
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol
F;1-36/Domain: signal sequence #status predicted <SIG>
F;35-260/Domain: PARP-like #status predicted <PARP>
F;512-527/Region: helical
F;512-1565/Product: collagen alpha 1(XI) chain #status predicted <MAT>
F;512-1542/Region: helical
F;1543-1555/Region: carboxyl-terminal nonhelical telopeptide
F;1566-1806/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;1583-1805/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;61-243,182-236/Disulfide bonds: #status predicted
F;505/Modified site: allysine (Lys) #status predicted
F;612,1452/Modified site: 5-hydroxylysine (Lys) #status predicted
F;612,1452/Binding site: carboxylate (Lys) (covalent) #status predicted

Query Match 86.2%; Score 930; DB 1; Length 1806;
Best Local Similarity 86.0%; Pred. No. 1.4e-50;
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy 1 GIRLKGKTKGKEDGPPGPKGDMGIRKDRGEIGPPGPRGEDGPEGPKGRGPNNDPGL 60
Db 769 GVRGLKSGKGEKEDGPPGPKGDMGLKGDREVGQVGRGEDGPEGPKGRAGPTGDPGP 828

Qy 61 GPPEKGLGVPLGPGVGRQPKSGTGFPGFPANKEKGRGTPGPKGPRGQGTGPR 120
Db 829 GQAGEKGLGVPLGPGVGRQPKSGTGFPGFPANKEKGRGTPGPKGPRGQGTGPR 888

Qy 121 GERPRGITGKPGKNSGGDGPAGPGRGPNQPGTGFPGPKGPPGPKGDXGLPGHP 180
Db 889 GSRGARGPTGKPGKGTSGGDPGPPGERGPPQGPVGFPGPKGPPGPKGDXGLPGHP 948

Qy 181 GQRGET 186
Db 949 GQRGET 954

RESULT 5
CGHUZE
collagen alpha 2(XI) chain precursor - human (fragment)
N;Alternate names: procollagen alpha 2(XI) chain
N;Contains: proline/arginine-rich protein (PARP)
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
C;Accession: S34790; A32645
R;Zhidkova, N.I.; Brewton, R.G.; Mayne, R.
FEBS Lett. 326, 25-28, 1993
A;Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage
chain.
A;Reference number: S34790; MUID:93314796; PMID:8325374
A;Accession: S34790
A;Molecule type: mRNA
A;Residues: 1-663 <ZHI>
A;Cross-references: EMBL:L18987; NID:G306439; PIDN:AAA35498.1; PID:G306440
R;Kimura, T.; Cheah, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M.;
J. Biol. Chem. 264, 13910-13916, 1989
A;Title: The human alpha2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and g
A;Reference number: A32645; MUID:89340485; PMID:2760050
A;Accession: A32645
A;Molecule type: DNA; mRNA
A;Residues: 586-1546 <KIM>
A;Cross-references: GB:J04974; NID:G180714; PIDN:AAA52034.1; PID:G180715
A;Note: parts of this sequence were determined by protein sequencing
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL11A2
A;Cross-references: GDB:I19788; OMIM:120290
A;Map position: 6p21.3-6p21.3

QY 1 GIRGLKGTGKEGEDGFPFGKGDWGI---KGRGETGPPGRGEDG-----PEGPKG 49
Db 1177 GPFGSKDGSGKEGVFFGLGASGPIGVYKGQGMFGPGQGLPTGPHVPEGPKG 1236
QY 50 RGGPNGD-----PGFLGPP-----GEKGLGVPLGYPGRQKPGKSGIGPFGP-- 92
Db 1237 DRGPQGPGLGHPGPMGPPGPPGNGPKGDKGNQWPGAPGVPGPKGDPGFGQMFGIGG 1296
QY 93 -PCANGEKGRGTTPGKGPGRGQRPPTGPRGERPRGITGKPGPKGNSGGDGPAGP----- 146
Db 1297 SPGITGSKGDMGLPGVPGFGQKGLPGLQGVKGDQGVPGPKLQGGPPGPPGYDVIK 1356
QY 147 --PGERGPNPGQPTGFPCKPKP-----PGPPGKDLGCHPQORGET 186
Db 1357 GEPGLPGPGLKGLQFPKPGKGOQGVTSVGLPGPVGFGDAPGQKGET 1410
RESULT 8
S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N;Alternate names: procollagen alpha 5(IV) chain
N;Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence revision 27-Feb-1997 #text change 21-Jul-2000
C;Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35
R;Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A;Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identi
n Alport syndrome patient.
A;Reference number: S22917; PMID:92316923; PMID:1352287
A;Accession: S22917
A;Molecule type: mRNA
A;Residues: 1-967 <ZH>
A;Cross-references: GB:M90464; NID:gl80826; PIDN:AAA52046.1; PID:G553234
R;Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A;Title: Structure of the human type IV collagen COL4A5 gene.
A;Reference number: A54365; PMID:94165049; PMID:8120014
A;Accession: A54365
A;Molecule type: DNA
A;Residues: 1-922 <ZH2>
A;Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAAC27816.1; PID
R;Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paepe, A.; Tryggva
Science 261, 1167-1169, 1993
A;Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited sm
A;Reference number: A57079; PMID:93361972; PMID:8356449
A;Accession: A57079
A;Molecule type: DNA
A;Residues: 1-27 <ZH4>
A;Cross-references: GB:Z37153; NID:G587203; PIDN:CAA85512.1; PID:G587204
R;Pihlajaniemi, T.; Pohjola, E.R.; Myers, J.C.
J. Biol. Chem. 265, 13758-13766, 1990
A;Title: Complete primary structure of the triple-helical region and the carboxyl-termin
A;Reference number: A37122; PMID:90337990; PMID:2380186
A;Accession: A37122
A;Molecule type: mRNA
A;Residues: 84-439, 'GS', 442-624, 'LALO', 629-666, 'ER', 669-887, 'R', 889-1264, 1271-1691 <PIH>
A;Cross-references: GB:U05558; EMBL:M58526; NID:g1314209
A;Note: submitted to the EMBL Data Library, February 1991
A;Note: the authors translated the codon GCC for residue 115 as Val
R;Renieri, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma
Hum. Mol. Genet. 1, 127-129, 1992
A;Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in
A;Reference number: I54317; PMID:93244772; PMID:1363780
A;Accession: I54317
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 313-324, 'E', 326-330 <REN>
A;Cross-references: GB:S59334; NID:g299946; PIDN:AAD13909.1; PID:G4261609
R;Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtyae, M.; Shows, T.B.; Tryggvason, K.
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
A;Title: Identification of a distinct type IV collagen alpha chain with restricted kidney

A;Reference number: A34850; PMID:90160375; PMID:1689491
A;Accession: A34850
A;Molecule type: mRNA
A;Residues: 914-1264, 1271-1691 <HOS>
A;Cross-references: EMBL:M31115; NID:gl80824; PIDN:AAA52045.1; PID:g180825
R;Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.
Genomics 9, 1-9, 1991
A;Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that
A;Reference number: A37969; PMID:91169491; PMID:2004755
A;Accession: S18850
A;Molecule type: DNA
A;Residues: 924-1264, 1271-1691 <ZH3>
A;Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMB
8; EMBL:M63470; EMBL:M63471; EMBL:M63473; NID:g177922; PIDN:AAA51558.1; PID:
R;Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; M
Kidney Int. 44, 1316-1321, 1993
A;Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex
A;Reference number: I56971; PMID:94133540; PMID:8301933
A;Accession: I56971
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1258-1276 <GUO1>
A;Cross-references: GB:S69168; NID:G545095; PIDN:AAAC06012.1; PID:G545096
A;Note: kidney splice form
A;Accession: I76598
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1284-1291, 'TFLGYLACLV' <GUO2>
A;Cross-references: GB:S69169; NID:G545097; PIDN:AAAC06013.1; PID:G545098
A;Note: frameshift mutation in patient with Alport syndrome
R;Myers, J.C.; Jones, T.A.; Pohjola, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; Sc
Am. J. Hum. Genet. 46, 1024-1033, 1990
A;Title: Molecular cloning of alphas(IV) collagen and assignment of the gene to the regi
A;Reference number: A35335; PMID:90252791; PMID:2339699
A;Accession: A35335
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1448-1477 <MYE>
R;Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Koitabaashi, Y.; Takada, T.; Yos
Kidney Int. 46, 1307-1314, 1994
A;Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord
A;Reference number: I56975; PMID:95156893; PMID:7853788
A;Accession: I56975
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1595-1602 <NAK>
A;Cross-references: GB:S75903; NID:g913882; PIDN:AAB33374.1; PID:g913883
A;Note: premature termination mutation from a patient with Alport syndrome; one other mu
R;Liemink, H.H.; Schroeder, C.H.; Brummer, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;
Genomics 17, 485-489, 1993
A;Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo
A;Reference number: I54188; PMID:94010948; PMID:8406498
A;Accession: I54188
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1604-1607, 'VHDAYKC' <LEM>
A;Cross-references: GB:S65767; NID:G425563; PIDN:AAD13967.1; PID:G4261667
A;Note: frameshift mutation from a patient with Alport syndrome; five other mutations a
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL4A5; ATS
A;Cross-references: GDB:120596; OMIM:303630
A;Map position: Xq22-Xq22
A;Introns: 97/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/
/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/3; 1152/1; 1185/1; 1
A;Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 5(
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-li
C;Function:
A;Description: minor structural component of extracellular basement membrane
C;Superfamily: collagen alpha 1(IV) chain

C;Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MATR
F;27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status p
F;27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F;42-1462/Region: interrupted helical
F;1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1473-1573/Domain: collagen IV carboxyl-terminal repeat <CTR>
F;1583-1687/Domain: collagen IV carboxyl-terminal repeat <CTR>
F;29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
F;15/Binding site: carbohydurate (Asn) (covalent) #status predicted
F;1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
F;1527-1533,1638-1644/Disulfide bonds: #status predicted
F;1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

Query Match 46.2%; Score 498.5; DB 1; Length 1691;
Best Local Similarity 46.0%; Pred. No. 6.6e-24;
Matches 108; Conservative 13; Mismatches 63; Indels 51; Gaps 7;

QY 1 GIRGLKTKGKEDGPPGFGDMGKIDRGE-----IGPPGRGEDGP--EGPKG 49
Db 1193 GPPGLGLSGOKDGGGLPGIPGNPLGPKGEFGFHGFGVQGGPPGPGSPGALGPKG 1252

QY 50 RCGPNGB-----PGLGPP-----GEKGLGVPLPGVP--GROGPKG 85
Db 1253 NPFGQPPGPRGTGFGGLGPPGPPGLPGNGIKGKNPQPGPLGLKGDQGGPPG 1312

QY 86 SIGFPPGPGANGKGRGTGPKPRGQRGTGPRGRGPR--GITGKPGP-----134
Db 1313 LQGNPCRPGLNMGKDPGLGVFGFGKMGKSPGVSAGPEGEPLGLGPPGPGPLGPPSG 1372

QY 135 -----KNSGGDGPAGPPGRRGNPGPGTGFPGPKGPPGPKGGLPHGPGRG 184
Db 1373 QSIIRKGDAGPPGTPGQGLKGLPGPGQGLGPTGTPGDRGNGLPGDGAGG 1427

RESULT 9
S28791
collagen alpha 1(XI) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
C;Accession: S28791
R;Nah, H.D.; Barenbaum, M.; Upholt, W.B.
J. Biol. Chem. 267, 22581-22586, 1992
A;Title: The chicken alpha(XI) collagen gene is widely expressed in embryonic tissues.
A;Reference number: S28791; MUID:93054557; PMID:1429607
A;Status: preliminary
A;Accession: S28791
A;Molecule type: mRNA
A;Residues: 1-888 <NAH>
A;Cross-references: EMBL:M88593; NID:g211619; PIDN:AAA48707.1; PID:g211620
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
F;665-887/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 46.1%; Score 497.5; DB 2; Length 888;
Best Local Similarity 48.8%; Pred. No. 4.5e-24;
Matches 103; Conservative 12; Mismatches 69; Indels 27; Gaps 6;

QY 1 GIRGLKTKGKEDGPPGFGDMGKIDRGEIGPPGRGEDGPEGKRGGP--NGDP 57
Db 10 GFGPGKPPGPKDGLPHGPGRGRTGTGFGTKTGPFGVGVGTGTGTGIGRGRHP 69

QY 58 GPLGPFGEKGLGV-----PGLFGYGRGQPKSGISGFP---GPPGANGS---KGR 102
Db 70 GPPGPPGEQQLPGAAGKEGAKGDPGQIPGKDPAGLRGFPGERGLPGAQGPAGLKGGE 129

QY 103 GTPGKPPGPRGQRGTGPRGRGRGTGPKPKGNSGGDGPAGPPGERGPNGB-----Q 156
Db 130 GPQGPFGVPSGPERGAAGTAGTIGLPGRPGQPPGPGPAGEKAGPEKGPQGPAGRDGVQ 189

QY 157 GPTGFGPKGPPGPKDGLP---GHPGQRG 184
Db 190 GPVGLPGPAGPSGPDGDKGIBGPGQKG 220

RESULT 10
S23297
collagen alpha 1(X) chain precursor - chicken
N;Alternate names: type X collagen
C;Species: Gallus gallus (chicken)
C;Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C;Accession: S23297; A31896; S65594; S77711; I50218
R;Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; M
maguchi, N.; Olsen, B.R.
in: Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114. Academic P
A;Title: The molecular biology of collagens with short triple-helical domains.
A;Reference number: S22243
A;Accession: S23297
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-674 <NIN>
R;LuValle, P.; Ninomiya, Y.; Rosenblum, N.D.; Olsen, B.R.
J. Biol. Chem. 263, 18378-18385, 1988
A;Title: The type X collagen gene. Introns sequences split the 5'-untranslated region an
A;Reference number: A31896; MUID:89054019; PMID:2461368
A;Accession: A31896
A;Molecule type: mRNA
A;Residues: 1-75 <LUV>
R;Ninomiya, Y.; Gordon, M.; van der Rest, M.; Schmid, T.; Linsemayer, T.; Olsen, B.R.
J. Biol. Chem. 261, 5041-5050, 1986
A;Title: The developmentally regulated type X collagen gene contains a long open readin
A;Reference number: I50218; MUID:86168227; PMID:3082876
A;Accession: S65594
A;Molecule type: DNA
A;Residues: 'T','9','D','11-12','EDMKLYLFTM','30-31','TCKSGRAFTTYMILQNMVADLVSSHT','48-89','L'
629,'PQAVLSLSWRTIKGSCQIQNMVSIPLNMIFLLSQVSVLLKNNIPLTMS' <NIN1>
A;Cross-references: EMBL:M13496; NID:g211699; PIDN:AAA48736.1; PID:g211700
A;Accession: S77711
A;Molecule type: protein
A;Residues: 104-112, 'X', 114-117, 453-466 <NIN2>
C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxyproline
F;1-18/Domain: signal sequence #status predicted <SIG>
F;547-673/Domain: complement C1q carboxyl-terminal homology <CIQ>
F;453,456/Modified site: hydroxyproline (Pro) #status experimental
F;611/Binding site: carbohydurate (Asn) (covalent) #status predicted

Query Match 45.8%; Score 494; DB 2; Length 674;
Best Local Similarity 46.2%; Pred. No. 5.9e-24;
Matches 102; Conservative 15; Mismatches 60; Indels 44; Gaps 6;

QY 1 GIRGLKTKGKEDGPPGFGDMGKIDRGEIGPPGRGEDGPEGPK-----GRGGPN 54
Db 295 GLFGMKGHRGSGPPGPPGPKGQDQGPAGVPGFPGPAGPQGNMGPQGLKGLPGENGLPGPK 354

QY 55 GPGGPLGP---PGEKGLGVPLG---PGVPGRGKSGISGFPFGPCANGKGGRGTPGKP 108
Db 355 GDMGVPVGPAGFPGAKGERGLGDLGKPGYPGSGGLPGPKGHPLGPKQGTGHAGPPGLP 414

QY 109 GP-----RQGRGTGP-----RGRGRGRTGTPKPKGKNSGD-----141
Db 415 GPGVPGVKGVPGINGEPGPGPSGIPGIRGTPGPGMPGAPGAKGAGAPGLPGPAGIA 474

QY 142 -----GPAGPPGERGPNQGPTGPPGPKGPPGPKGKGLP 177
Db 475 TKGLRGPMGPPGPPGPKGNSGEPGLPGPPGPPGPPGQSTIP 515

RESULT 11
CGH2V
collagen alpha 2(V) chain precursor - human
C;Species: Homo sapiens (man)
C;Date: 31-Jul-1989 #sequence_revision 28-Jul-1995 #text_change 31-Dec-2000
C;Accession: A31427; A54555; S43643; A35874; I55239; I59025; A25374; A30017
R;Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989

A;Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the structure of alpha2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the length, is formed with desmosine cross-links made from lysine and allysine residues

A;Gene: GDB:COL5A2

A;Cross-references: GDB:119064; OMIM:120190

A;Map position: 2q31-2q31

A;Introns: 33/1; 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1448/3; 1448/3; 1448/3

A;Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHUIV), a alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the length, is formed with desmosine cross-links made from lysine and allysine residues

C;Function:

A;Description: structural component of extracellular fibrous polymer associated with cell growth and differentiation

A;Note: may play a role in controlling the lateral growth of collagen I fibrils

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline; F;1-26/Domain: signal sequence #status predicted <Sig>

F;1-26/Domain: signal sequence #status predicted <Sig>

F;27-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>

F;27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>

F;27-108/Region: nonhelical

F;27-99/Domain: von Willebrand factor type C repeat homology <VWC>

F;40-99/Region: helical

F;187-208/Region: helical

F;187-208/Region: nonhelical

F;209-1225/Region: helical

F;503-505/Region: cell attachment (R-G-D) motif

F;941-943/Region: cell attachment (R-G-D) motif

F;1064-1066/Region: cell attachment (R-G-D) motif

F;1067-1069/Region: cell attachment (R-G-D) motif

F;1097-1099/Region: cell attachment (R-G-D) motif

F;1124-1126/Region: cell attachment (R-G-D) motif

F;1133-1135/Region: cell attachment (R-G-D) motif

F;1225-1250/Region: carboxyl-terminal nonhelical telopeptide

F;1251-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>

F;1251-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

F;127/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F;193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted

F;194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F;201/Modified site: allysine (Lys) #status predicted

F;290, 293, 296, 608, 614, 1004, 1007, 1013, 1028, 1034/Modified site: 4-hydroxyproline (Pro) #status predicted

F;299, 1139/Binding site: 5-hydroxylysine (Lys) #status predicted

F;299, 1139/Binding site: carboxylate (Lys) (covalent) #status predicted

F;1025/Modified site: 5-hydroxylysine (Lys) #status predicted

F;1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted

F;1259, 1397/Binding site: carboxylate (Asn) (covalent) #status predicted

F;1293, 1299, 1325/Disulfide bonds: interchain #status predicted

F;1333-1494, 1402-1447/Disulfide bonds: #status predicted

Query Match 45.5%; Score 491; DB 1; Length 1496;

Best Local Similarity 51.4%; Pred. No. 1.7e-23;

Matches 95; Conservative 12; Mismatches 78; Indels 0; Gaps 0;

QY	1	GIRGLKGTGKGGEDGFPFGKDMGKIGDRGEIGPPGRGSDGPGKGRGPNGDGPGPL	60
DB	291	GAPGLPGLKGRHGHKLEGGKGEVGAPOGSKGAGTGPAGMGPLGPRGMPGERGLGPGQ	350
QY	61	GPPGSEKGLGVPLGYPGRQPKGSIGFPGFPGANGKGGRTGPGKPGRGQRTGPR	120
DB	351	GAPGQGAHGMFGKGPWPLGIPGSSGFPNGPMKGAGTGAAGPQGQRTGPP	410
QY	121	GERGPRGITGKPGKNGSGDGPAGPGERGNGPQGTGPGKPGPPKDGKLGCHP	180
DB	411	GVGSGPLFGAIGTDGTPGKPTSGPTSGPSAGPFGSPGQSGTGPQNSGLGDP	470
QY	181	GORGE 185	
DB	471	GFKE 475	

RESULT 12

S59856

collagen alpha 1(III) chain precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 13-Aug-1999

C;Accession: S59856; S62120; S16373

R;Toman, P.D.; de Crombrughe, B.

Gene 147, 161-168, 1994

A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA

A;Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the structure of alpha2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the length, is formed with desmosine cross-links made from lysine and allysine residues

A;Gene: GDB:COL5A2

A;Cross-references: GDB:119064; OMIM:120190

A;Map position: 2q31-2q31

A;Introns: 33/1; 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1448/3; 1448/3; 1448/3

A;Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHUIV), a alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the length, is formed with desmosine cross-links made from lysine and allysine residues

C;Function:

A;Description: structural component of extracellular fibrous polymer associated with cell growth and differentiation

A;Note: may play a role in controlling the lateral growth of collagen I fibrils

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline; F;1-26/Domain: signal sequence #status predicted <Sig>

F;1-26/Domain: signal sequence #status predicted <Sig>

F;27-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>

F;27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>

F;27-108/Region: nonhelical

F;27-99/Domain: von Willebrand factor type C repeat homology <VWC>

F;40-99/Region: helical

F;187-208/Region: helical

F;187-208/Region: nonhelical

F;209-1225/Region: helical

F;503-505/Region: cell attachment (R-G-D) motif

F;941-943/Region: cell attachment (R-G-D) motif

F;1064-1066/Region: cell attachment (R-G-D) motif

F;1067-1069/Region: cell attachment (R-G-D) motif

F;1097-1099/Region: cell attachment (R-G-D) motif

F;1124-1126/Region: cell attachment (R-G-D) motif

F;1133-1135/Region: cell attachment (R-G-D) motif

F;1225-1250/Region: carboxyl-terminal nonhelical telopeptide

F;1251-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>

F;1251-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

F;127/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F;193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted

F;194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F;201/Modified site: allysine (Lys) #status predicted

F;290, 293, 296, 608, 614, 1004, 1007, 1013, 1028, 1034/Modified site: 4-hydroxyproline (Pro) #status predicted

F;299, 1139/Binding site: 5-hydroxylysine (Lys) #status predicted

F;299, 1139/Binding site: carboxylate (Lys) (covalent) #status predicted

F;1025/Modified site: 5-hydroxylysine (Lys) #status predicted

F;1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted

F;1259, 1397/Binding site: carboxylate (Asn) (covalent) #status predicted

F;1293, 1299, 1325/Disulfide bonds: interchain #status predicted

F;1333-1494, 1402-1447/Disulfide bonds: #status predicted

Query Match 45.5%; Score 491; DB 1; Length 1496;

Best Local Similarity 51.4%; Pred. No. 1.7e-23;

Matches 95; Conservative 12; Mismatches 78; Indels 0; Gaps 0;

QY	1	GIRGLKGTGKGGEDGFPFGKDMGKIGDRGEIGPPGRGSDGPGKGRGPNGDGPGPL	60
DB	291	GAPGLPGLKGRHGHKLEGGKGEVGAPOGSKGAGTGPAGMGPLGPRGMPGERGLGPGQ	350
QY	61	GPPGSEKGLGVPLGYPGRQPKGSIGFPGFPGANGKGGRTGPGKPGRGQRTGPR	120
DB	351	GAPGQGAHGMFGKGPWPLGIPGSSGFPNGPMKGAGTGAAGPQGQRTGPP	410
QY	121	GERGPRGITGKPGKNGSGDGPAGPGERGNGPQGTGPGKPGPPKDGKLGCHP	180
DB	411	GVGSGPLFGAIGTDGTPGKPTSGPTSGPSAGPFGSPGQSGTGPQNSGLGDP	470
QY	181	GORGE 185	
DB	471	GFKE 475	

RESULT 12

S59856

collagen alpha 1(III) chain precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 13-Aug-1999

C;Accession: S59856; S62120; S16373

R;Toman, P.D.; de Crombrughe, B.

Gene 147, 161-168, 1994

A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA

A;Reference number: S59856; MUID:95011609; PMID:7926795
A;Accession: S59856
A;Molecule type: DNA
A;Residues: 1-1464 <TOM>
A;Cross-references: EMBL:X52046
R;Toman, D.
submitted to the EMBL Data Library, November 1994
A;Reference number: S62120
A;Accession: S62120
A;Molecule type: DNA
A;Residues: 1-866, 'G', 868-1464 <TOM>
A;Cross-references: EMBL:X52046; NID:9575321; PIDN:CAA36279.1; PID:9575322
R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A;Reference number: S16176; MUID:91274355; PMID:2054384
A;Accession: S16173
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1464 <MET>
A;Cross-references: EMBL:X57983; NID:950476; PIDN:CAA1048.1; PID:950477
C;Genetics:
A;Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 295/3; 673/3; 706/3; 742/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-154/Domain: propeptide #status predicted <PRO>
F;32-92/Domain: von Willebrand factor type C repeat homology <VMC>
F;155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <PCC>
Query Match 45.4%; Score 489.5; DB 2; Length 1464;
Best Local Similarity 46.2%; Pred. No. 2.1e-23;
Matches 103; Conservative 17; Mismatches 64; Indels 39; Gaps 7;
QY 1 GIRGLKTKGKEDGFPFGKDMGK---GDRGEI GPPGRGE-----DGPE 45
DB 698 GPPGEGGKGPAGPPGPPGASGSLQGMFGRGPGSGPKGEPGAGADGVGKD 757
QY 46 GPKGRGPNPDGPLGPGKGLGVPLGYP-----GROGPKSGISGPPPGAN 96
DB 758 GPRGPA GIPGPPAGAGGKGGGSGPLGFIAGPRGPGRGEHGGPPGAGPPGAGQQN 817
QY 97 GE---KGRGTTPGK-----PGPRGQPTGPRGPRGITGK---PGPKGNS---GGD 141
DB 818 GEPKAGKGERGAPGKGGGPPGAGPTGSSGPPGPPGQGVKGRGSPGPTAGPFGGR 877
QY 142 GPAGPPGPRGPNQGGTGTGPPGKGGPPGPKDGLPHQPGQR 184
DB 878 GLPGPPGNNNGPPGSPGAPGKDGPPGAGNSGPNPAG 920
RESULT 13
A32856
collagen alpha 1(VI) chain precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 14-Jul-2003
C;Accession: A32856; I50587; I50587; I50586
R;Bonaldo, P.; Russo, V.; Bucciolotti, F.; Bressan, G.M.; Colombatti, A.
J. Biol. Chem. 264, 5575-5580, 1989
A;Title: Alpha-1 chain of chick type VI collagen. The complete cDNA sequence reveals a H
A;Reference number: A32856; MUID:89174602; PMID:2784434
A;Accession: A32856
A;Molecule type: mRNA
A;Residues: 1-1019 <BON>
A;Cross-references: GB:J04598; NID:9576463; PIDN:AAB59954.1; PID:g2111354
A;Note: 479-Asn and 620-Asn was also found
R;Walchli, C.; Koller, E.; Trueb, B.
Eur. J. Biochem. 205, 583-589, 1992
A;Title: Structural comparison of the chicken genes for alpha 1(VI) and alpha 2(VI) coll
A;Reference number: I50587; MUID:92241293; PMID:1572359
A;Accession: I50587

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1019 <WAL>
A;Cross-references: EMBL:X57998; NID:962874; PIDN:CAA41062.1; PID:962875
A;Accession: I50627
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1019 <WA2>
A;Cross-references: EMBL:X64458; NID:963301; PIDN:CAA45788.1; PID:963302
R;Koller, E.; Trueb, B.
Eur. J. Biochem. 208, 769-774, 1992
A;Title: Characterization of the chicken alpha 1(VI) collagen promoter.
A;Reference number: I50586; MUID:93011107; PMID:1396681
A;Accession: I50586
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-75 <KOL>
A;Cross-references: EMBL:X57987; NID:962872; PIDN:CAA41053.1; PID:962873
C;Genetics:
A;Gene: Col6A1
A;Introns: 33/1; 76/2; 143/2; 196/3; 237/3; 244/3; 251/3; 266/3; 284/3; 299/3; 308/3; 3
06/1; 650/3; 687/2; 748/3; 807/1; 817/1
C;Superfamily: collagen VI; von Willebrand factor type A repeat homology
C;Keywords: cell binding; coiled coil; disulfide bond; extracellular matrix; glycoprote
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1019/Product: collagen alpha 1(VI) chain #status predicted <MAT>
F;35-213/Domain: von Willebrand factor type A repeat homology <VWA1>
F;255-590/Domain: collagenous #status predicted <COL>
F;440-442/Region: cell attachment (R-G-D) motif
F;476-478/Region: cell attachment (R-G-D) motif
F;529-531/Region: cell attachment (R-G-D) motif
F;611-788/Domain: von Willebrand factor type A repeat homology <VWA2>
F;822-989/Domain: von Willebrand factor type A repeat homology <VWA3>
F;212,514,535,666,799,887/Binding site: carbohydrate (Asn) #status predicted
Query Match 45.3%; Score 488.5; DB 1; Length 1019;
Best Local Similarity 46.7%; Pred. No. 1.8e-23;
Matches 99; Conservative 24; Mismatches 62; Indels 27; Gaps 4;
QY 1 GIRGLKTKGKEDGFPFGKDM-----GIRKDRGEI GPPGRGDPGPKGRG 51
DB 297 GKKGDKGRGKGRGKAKGKRGKRGIDGDMKEAGYGLPCKGSGPFGTQGP 356
QY 52 GPNQDPGPIGP-----PGEKGLGVPLGYPGPRGPKSGISGIFPFGANGKGGRTG 105
DB 357 GPKGDPGAYGPKGKGPEDGPKRGQIGPSGKAGPGRGEPGLGEGSPGAD 416
QY 106 GRKPPR---GORGPTGPRGPRGITGKPGKNSGGDPGPRGPRGPNQGTGF- 161
DB 417 GPFGRSGNGRGPSPGDRGRDLGEPFGPDGREGPLGPDGQSGPPGPKGYR 476
QY 162 -----PPKPGPPGPKGKGLGHPHGORGE 185
DB 477 GDDGPRNGPKSPGAPGLPGDGLMGERGE 508
RESULT 14
A40333
collagen alpha 1'(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 16-Jul-1999
C;Accession: A40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis e
A;Reference number: A40333; MUID:92011899; PMID:1918153
A;Accession: A40333
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1492 <SUA>
A;Cross-references: GB:M63596
A;Note: this sequence is presented as substitutions relative to another sequence in a f
es they replace; the appropriate interpretation of the sequence figure was reconstructed.

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VMC>
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

Query Match 45.3%; Score 488.5; DB 2; Length 1492;
Best Local Similarity 41.5%; Pred.No.2.5e-23;
Matches 103; Conservative 18; Mismatches 64; Indels 63; Gaps 4;

Qy 1 GIRGLKTKGEKGEDGPFPGFKGDMMGIKGRGEIGPPRGEDGPEGPKRGGPNNGDP--- 57
Db :
446 GPQATGPLGPKGTGPDGTAGFKGHGPKGEISAGPQGAPPAGEEGKKRGARGEPGA 505
Qy :
58 GPLGPPGEKGLGVPLGYVPGRQPKGSIGFPFFPGANGKEKGGRTGTPGKP----- 108
Db :
506 GPLGPPGERGAPGNRGFPDQLAGPKGAPGERGVPLGLGPKGGNDPRGPERGPLPGAR 565
Qy :
109 -----GPRQGRGPT-----GPRERPRGITKPKGPKGNS----- 138
Db :
566 GLTGRPDGADPGQRKVGPSGAAGEDGRPPQPQARGQPGVMGFPPGPKGAPFGKAGEK 625
Qy :
139 -----GGDGPAHPCEERGPNPGPQQGTGTFPGPKGPPGPKDKGLP 177
Db :
626 GLGAPGLRLPLGKDGTGAGCPNPGAPAGEREQGPPGPGFQGLPPSPSGEGGKP 685
Qy :
178 GHPCQORGE 185
Db :
686 GDQGVPGE 693

RESULT 15
CGHU7L
collagen alpha 1(III) chain precursor - human
N:Alternate names: procollagen alpha 1(III) chain
C:Species: Homo sapiens (man)
C>Date: 24-Apr-1984 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
C:Accession: S05272; S04642; P80011; S01726; S04887; A90399; A94562; I51868; S59511; A90399 submitted to the EMBL Data Library, February 1989
R:Prockop, D.J.
A:Reference number: S05272
A:Accession: S05272
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1240,'v',1242-1466 <PCC>
A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
R:Ala-kokko, L.; Kontusari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A>Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human
enferes.

A:Reference number: S04642; MUID:89350838; PMID:2764886
A:Accession: S04642
A:Molecule type: mRNA
A:Residues: 1-1196 <ALA>
A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
A>Note: the complete sequence is not shown
R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A>Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene
A:Reference number: PE0011; MUID:89378752; PMID:2777083
A:Accession: PE0011
A:Molecule type: DNA
A:Residues: 1-176 <BNB>
A:Cross-references: GB:M29339; NID:g180813; PIDN:AA52040.1; PID:g180814
R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
Nucleic Acids Res. 16, 7201, 1988
A>Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A:Reference number: S01726; MUID:88303360; PMID:3405773
A:Accession: S01726
A:Molecule type: mRNA
A:Residues: 1-170 <TON>
A:Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061
A>Note: The authors translated the codon CAG for residue 154 as His
R:Janeczko, R.A.; Ramirez, F.

Nucleic Acids Res. 17, 6742, 1999

A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen
A:Reference number: S04887; MUID:89386015; PMID:2780304
A:Accession: S04887
A:Molecule type: mRNA
A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634, 'A'
A:Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA3387.1; PID:G930045
A:Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R:Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A:Reference number: A90399; MUID:77134724; PMID:557335
A:Accession: A90399
A:Molecule type: protein
A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294, 'S', 296-398 <SEY1>
A:Experimental source: liver
A:Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose
R:Seyer, J.M.
submitted to the Atlas, December 1977
A:Reference number: A94562
A:Accession: A94562
A:Molecule type: protein
A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A:Experimental source: liver
A:Note: author submitted corrections to A90399
R:Miliewicz, D.M.; Wit, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A:Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
isping.
A:Reference number: I51868; MUID:93304430; PMID:8317500
A:Accession: I51868
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 186-194 <MIL>
A:Cross-references: GB:S62925; NID:G386425; PIDN:ADI3937.1; PID:G4261637
R:Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A:Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1
A:Reference number: S59511; MUID:96067614; PMID:7487954
A:Accession: S59511
A:Molecule type: mRNA
A:Residues: 302-423 <CHI>
A:Cross-references: GB:S79877; NID:G1195576; PIDN:AAB35615.1; PID:G1195577
R:Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A:Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides
A:Reference number: A90414; MUID:79000343; PMID:687591
A:Accession: A90414
A:Molecule type: protein
A:Residues: 399-675, 'N', 677-727 <SEY3>
A:Experimental source: liver
R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A:Title: G to T transversion at position +5 of a splice donor site causes skipping of th
A:Reference number: I55349; MUID:91161621; PMID:1672129
A:Accession: I55349
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 537-605 <LEE>
A:Cross-references: GB:MS9312; NID:G180815; PIDN:AAAS2041.1; PID:G180816
R:Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty
A:Reference number: A90438; MUID:80198282; PMID:6246925
A:Accession: A90438
A:Molecule type: protein
A:Residues: 728-895, 'A', 897-964 <SEY4>
A:Experimental source: liver
R:Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
J. Biol. Chem. 265, 17070-17077, 1990
A:Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
A:Reference number: A38303; MUID:91009133; PMID:2145268
A:Accession: A38303

A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA859383.1; PID:9
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn
R;Mankoo, B.S.; Dalglish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018, 'S', 1020-1183, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-
A;Experimental source: liver
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9183-9194, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage
A;Reference number: A93551; MUID:85087944; PMID:6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155, 'P', 1157-1466 <LOI>
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1
R;Miskulin, M.; Dalglish, R.; Kluge-Becker, B.; Renard, S.E.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: I52393; MUID:86187804; PMID:3754462
A;Accession: I52393
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MS>
A;Cross-references: GB:M13146; NID:G180415; PIDN:AA52003.1; PID:G180416
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I79359
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EWA>
A;Cross-references: GB:M11134; NID:G180417; PIDN:AA52004.1; PID:G180418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. F
A;Reference number: A92516; MUID:85157600; PMID:2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given d
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C
C;Genetics:
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
er of their length, is formed with desmosine cross-links made from lysine and alllysine x
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains inte
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
F;1-23/Domain: signal sequence #status predicted <SIG>

F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>
F;31-91/Domain: von Willebrand factor type C repeat homology <VWC>
F;154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;154-167/Region: amino-terminal nonhelical telopeptide
F;168-1196/Region: helical
F;1091-1093/Region: cell attachment (R-G-D) motif
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F;154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;161,1212/Modified site: allysine (Lys) #status predicted
F;263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
F;263/Binding site: carboxylate (Lys) (covalent) #status experimental
F;584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F;948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
F;1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 45.1%; Score 486.5; DB 1; Length 1466;
Best Local Similarity 46.4%; Pred. No. 3.2e-23;
Matches 98; Conservative 13; Mismatches 73; Indels 27; Gaps 3;
QY 1 GTRGLKGTGKEKEDGFPKGDNGI---KGDREITGPPCPRGEDGEGPKGKGNGDP 57
Db 711 GPFPPGAAATPGLQMGPGERGGLGSPGPKGDKGEPGPGADGVPKDGPRGTPTGPP 770
QY 58 GPLGPPGEGKLGVPGLPGVP-----GROGPKGSI GFPGFPGANGKGGRTGPKP 108
Db 771 GPAGQPDKGEGGAPGLGIAGPRGSPGERGTGPPGPPAGFPAGQNGEPGKGERGAP 830
QY 109 GPRGQ-----RGPTGPRGERGPRGITGKPGPKGNSGGDGSPAGPPGGERGN 153
Db 831 GEKGGGPPGVPAGPPGSGSPAGPPGQGVKGERGSPGPGAAAGFPGARGLPFGPSNGNP 890
QY 154 GPOGPTGFPKPGKPPGPPKDGKGLPGHPGQRG 184
Db 891 GPFPSGSPKDGPPGPPGAGNTGAPSGPSVSG 921

Search completed: May 3, 2004, 13:11:49
Job time : 9.09953 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:01:53 ; Search time 6.07465 Seconds
(without alignments)
1594.339 Million cell updates/sec

Title: US-09-775-964-6

Perfect score: 1079

Sequence: 1 GIRLKGTYGKEDGFGP.....PPGPPKQGLPGHQGRGET 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1079	100.0	1838	1	CA15_HUMAN
2	953	88.3	911	1	CA18_BOVIN
3	953	88.3	1806	1	CA18_HUMAN
4	945	87.6	1804	1	CA1B_MOUSE
5	907	84.1	1736	1	CA2B_HUMAN
6	901	83.5	1736	1	CA2B_MOUSE
7	737	68.3	1745	1	CA35_HUMAN
8	509.5	47.2	1049	1	CA13_BOVIN
9	501.5	46.5	1685	1	CA54_HUMAN
10	501	46.4	1669	1	CA14_MOUSE
11	496	46.0	674	1	CA1A_CHICK
12	493.5	45.7	1262	1	CA13_CHICK
13	491	45.5	1496	1	CA25_HUMAN
14	488.5	45.3	1019	1	CA16_CHICK
15	487.5	45.2	1464	1	CA13_MOUSE
16	487	45.1	1366	1	CA21_HUMAN
17	486.5	45.1	1466	1	CA13_HUMAN
18	486	45.0	1669	1	CA1A_HUMAN
19	485.5	45.0	674	1	CA1A_BOVIN
20	480	44.5	1372	1	CA21_RAT
21	479.5	44.4	471	1	CA34_BOVIN
22	479.5	44.4	671	1	CA11_RAT
23	479.5	44.4	1516	1	CA1H_HUMAN
24	478.5	44.3	1464	1	CA11_HUMAN
25	477.5	44.3	680	1	CA1A_HUMAN
26	477.5	44.3	1453	1	CA11_MOUSE
27	477	44.2	747	1	CA12_BOVIN
28	477	44.2	1690	1	CA44_HUMAN
29	476.5	44.2	1418	1	CA12_HUMAN
30	476.5	44.2	1460	1	CA11_CANFA
31	476	44.1	779	1	CA11_BOVIN
32	476	44.1	1366	1	CA21_CANFA
33	475.5	44.1	1355	1	CA21_RANCA

RESULT 1				
ID	CA15_HUMAN	STANDARD;	PRT;	1838 AA.
AC	P20908;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Collagen alpha 1(V) chain precursor.			
GN	COL5A1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.			
RX	MEDLINE=91302336; PubMed=2071595;			
RA	Takahara K., Seto Y., Okasawa K., Okamoto N., Noda A., Yaoi Y.,			
RA	Kato I.;			
RT	"Complete primary structure of human collagen alpha 1 (V) chain.";			
RL	J. Biol. Chem. 266:13124-13129(1991).			
RN	[2]			
RP	SEQUENCE OF 621-822.			
RC	TISSUE=Chorioamniotic membrane;			
RX	MEDLINE=89227189; PubMed=2496661;			
RA	Seyer J.M., Kang A.H.;			
RT	"Covalent structure of collagen: amino acid sequence of three			
RT	cyanoen bromide-derived peptides from human alpha 1(V) collagen			
RT	chain.";			
RL	Arch. Biochem. Biophys. 271:120-129(1989).			
RN	[3]			
RP	SEQUENCE OF 823-950, AND HEPARIN-BINDING.			
RX	MEDLINE=90366601; PubMed=2203476;			
RA	Yaoi Y., Hashimoto K., Koitabashi H., Takahara K., Ito M., Kato I.;			
RT	"Primary structure of the heparin-binding site of type V collagen.";			
RL	Biochim. Biophys. Acta 1035:139-145(1990).			
RN	[4]			
RP	SEQUENCE OF 556-571.			
RC	TISSUE=Placenta;			
RX	MEDLINE=92239022; PubMed=1571108;			
RA	Mann K.;			
RT	"Isolation of the alpha 3-chain of human type V collagen and			
RT	characterization by partial sequencing";			
RL	Biol. Chem. Hoppe-Seyler 373:69-75(1992).			
RN	[5]			
RP	SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.			
RC	TISSUE=Chorioamniotic membrane;			
RX	MEDLINE=94237164; PubMed=8181482;			
RA	Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champliand M.P.,			
RA	Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;			
RT	"Diversity in the processing events at the N-terminus of type-V			
RT	collagen.";			
RL	Eur. J. Biochem. 221:987-995(1994).			
RN	[6]			
RP	DISEASE, AND VARIANT EDS-I SER-1639.			
RX	MEDLINE=97195540; PubMed=9042913;			
RA	de Paeppe A., Nuytinck L., Hauser I., Anton-Lamprecht I.,			

and reveals that the expression of the gene is not restricted to cartilaginous tissue.";
J. Biol. Chem. 263:17159-17166(1988).
[4]
ALTERNATIVE SPLICING.
TISSUE=Blood;
MEDLINE=95238468; PubMed=7721876;
Zhukova N.I., Justice S.K., Mayne R.;
RT "Alternative mRNA processing occurs in the variable region of the
pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";
J. Biol. Chem. 270:9486-9493(1995).
[5]
VARIANT STL2 VAL-625.
MEDLINE=97026296; PubMed=8872475;
Richards A.J., Yates J.R.W., Williams R., Payne S.J., Pope F.M.,
RA Scott J.D., Sneed M.P.;
RT "A family with Stickler syndrome type 2 has a mutation in the COL11A1
gene resulting in the substitution of glycine 97 by valine in
alpha-1(XI) collagen.";
Hum. Mol. Genet. 5:1339-1343(1996).
CC -I- FUNCTION: May play an important role in fibrillogenesis by
controlling lateral growth of collagen II fibrils.
CC -I- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),
alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational
modification of alpha 1(XI). Alpha 1(XI) can also be found instead
of alpha 3(XI)=1(II).
CC -I- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=3;
Comment=Additional isoforms seem to exist. There is alternative
usage of exon IIA or exon IIB. Transcripts containing exon IIA
or IIB are present in cartilage, but exon IIB is preferentially
utilized in transcripts from tendon;
Name=A;
IsoId=P12107-1; Sequence=Displayed;
Name=B;
IsoId=P12107-2; Sequence=VSP_001145;
Name=C;
IsoId=P12107-3; Sequence=VSP_001146;
CC -I- TISSUE SPECIFICITY: Cartilage, placenta and some tumor or virally
transformed cell lines. Isoforms using exon IIA or IIB are found
in the cartilage while isoforms using only exon IIB are found in
the tendon.
CC -I- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -I- DISEASE: Defects in COL11A1 are the cause of Stickler syndrome
type 2 (STL2) [MIM:604841]; also known as Stickler syndrome
vitreous type 2, or beaded vitreous type, due to the presence of
irregularly thickened fiber bundles throughout vitreous cavity.
CC Stickler syndrome (hereditary progressive arthro-ophthalmopathy)
is an autosomal dominant disorder characterized by progressive
myopia beginning in the first decade of life, vitreo-retinal
degeneration, retinal detachment, cleft palate, midfacial
hypoplasia, osteoarthritis, and sensorineural hearing loss.
CC -I- DISEASE: Defects in COL11A1 are the cause of Marshall syndrome
[MIM:154780], an autosomal dominant disorder with ocular, oro-
facial, auditory and skeletal manifestations. It shares several
features with Stickler syndrome, such as midfacial hypoplasia,
high myopia, and sensorineural-hearing deficit.
CC -I- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -I- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.
CC -I- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

DR EMBL; J04177; AA51891.1; -;
DR EMBL; AF101112; AAF04724.1; -;
DR EMBL; AF101079; AAF04724.1; JOINED.

DR EMBL; AF101080; AAF04724.1; JOINED.
DR EMBL; AF101081; AAF04724.1; JOINED.
DR EMBL; AF101082; AAF04724.1; JOINED.
DR EMBL; AF101083; AAF04724.1; JOINED.
DR EMBL; AF101084; AAF04724.1; JOINED.
DR EMBL; AF101085; AAF04724.1; JOINED.
DR EMBL; AF101086; AAF04724.1; JOINED.
DR EMBL; AF101087; AAF04724.1; JOINED.
DR EMBL; AF101088; AAF04724.1; JOINED.
DR EMBL; AF101089; AAF04724.1; JOINED.
DR EMBL; AF101090; AAF04724.1; JOINED.
DR EMBL; AF101091; AAF04724.1; JOINED.
DR EMBL; AF101092; AAF04724.1; JOINED.
DR EMBL; AF101093; AAF04724.1; JOINED.
DR EMBL; AF101094; AAF04724.1; JOINED.
DR EMBL; AF101095; AAF04724.1; JOINED.
DR EMBL; AF101096; AAF04724.1; JOINED.
DR EMBL; AF101097; AAF04724.1; JOINED.
DR EMBL; AF101098; AAF04724.1; JOINED.
DR EMBL; AF101099; AAF04724.1; JOINED.
DR EMBL; AF101100; AAF04724.1; JOINED.
DR EMBL; AF101101; AAF04724.1; JOINED.
DR EMBL; AF101102; AAF04724.1; JOINED.
DR EMBL; AF101103; AAF04724.1; JOINED.
DR EMBL; AF101104; AAF04724.1; JOINED.
DR EMBL; AF101105; AAF04724.1; JOINED.
DR EMBL; AF101106; AAF04724.1; JOINED.
DR EMBL; AF101107; AAF04724.1; JOINED.
DR EMBL; AF101108; AAF04724.1; JOINED.
DR EMBL; AF101109; AAF04724.1; JOINED.
DR EMBL; AF101110; AAF04724.1; JOINED.
DR EMBL; AF101111; AAF04724.1; JOINED.
DR EMBL; AF101112; AAF04724.1; JOINED.
DR EMBL; AF101079; AAF04726.1; JOINED.
DR EMBL; AF101080; AAF04726.1; JOINED.
DR EMBL; AF101081; AAF04726.1; JOINED.
DR EMBL; AF101082; AAF04726.1; JOINED.
DR EMBL; AF101083; AAF04726.1; JOINED.
DR EMBL; AF101084; AAF04726.1; JOINED.

```

DR EMBL; AF101085; AAF04726.1; JOINED.
DR EMBL; AF101086; AAF04726.1; JOINED.
DR EMBL; AF101087; AAF04726.1; JOINED.
DR EMBL; AF101088; AAF04726.1; JOINED.
DR EMBL; AF101089; AAF04726.1; JOINED.
DR EMBL; AF101090; AAF04726.1; JOINED.
DR EMBL; AF101091; AAF04726.1; JOINED.
DR EMBL; AF101092; AAF04726.1; JOINED.
DR EMBL; AF101093; AAF04726.1; JOINED.
DR EMBL; AF101094; AAF04726.1; JOINED.
DR EMBL; AF101095; AAF04726.1; JOINED.
DR EMBL; AF101096; AAF04726.1; JOINED.
DR EMBL; AF101097; AAF04726.1; JOINED.
DR EMBL; AF101098; AAF04726.1; JOINED.

Query Match      88.3%; Score 953; DB 1; Length 1806;
Best Local Similarity 88.2%; Pred. No. 6.3e-48;
Matches 164; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

Qy 1 GIRLKGKGEKGDGPGFGKDMGIKDRGETGPPGPRGDEGPGKGGNGDGPPL 60
Db 769 GVRGLKSGKEGDEGDFGFGKDMGLKDRGEVQIGPRGEDGPGKGRAGTDPGPGS 828

Qy 61 GPGKEKGLGVGLPGVPGRQKSGSIFGFGPGANGKEKGRGTPKPGPRGQGTGPR 120
Db 829 GQAGEKGLGVGLPGVPGRQKSGSTGFGFGANGKEKGRGTPKPGPRGQGTGPR 888

Qy 121 GERGPRGITKPGPKNGSGDGGAGPPGPRGPNQPGTGFPGPKGPPGKDGLPGH 180
Db 889 GSRGARGTKPGPKGTSGDGGPPGPRGPNQPGTGFPGPKGPPGKDGLPGH 948

Qy 181 GQGET 186
Db 949 GQGET 954

RESULT 4
ID -CAIB MOUSE STANDARD; PRT; 1804 AA.
AC Q61245; Q64047;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(XI) chain precursor.
GN COL1A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RN TISSUE=Embryo;
RX MEDLINE=96015067; PubMed=8530046;
RA Yoshioka H., Inoguchi K., Khaleduzzaman M., Ninomiya Y.,
RA Andrikopoulos K., Ramirez F.;
RT "Coding sequence and alternative splicing of the mouse alpha 1(XI)
RL collagen gene (Col1a1).";
RL Genomics 28:337-340(1995).
RN [2]
RP SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.
RC STRAIN=C57BL/6;
RX MEDLINE=95163095; PubMed=7859283;
RA Li Y., Lacorda D.A., Warman M.L., Beier D.R., Yoshioka H.,
RA Ninomiya Y., Oxford J.T., Morris N.P., Andrikopoulos K.,
RA Ramirez F., Wardell B.B., Lifferth G.D., Teuscher C., Woodward S.R.,
RA Taylor B.A., Seegmiller R.E., Olsen B.R.;
RT "A fibrillar collagen gene, Col1a1, is essential for skeletal
RT morphogenesis.";
RL Cell 80:423-430(1995).
CC -i- FUNCTION: May play an important role in fibrillogenesis by
CC controlling lateral growth of collagen II fibrils.
CC -i- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),
CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational

```

```

CC modification of alpha 1(XI). Alpha 1(V) can also be found instead
CC of alpha 3(XI)=1(XI) (By similarity).
CC -i- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q61245-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q61245-2; Sequence=VSP_001147;
CC -i- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -i- DISEASE: Defects in COL1A1 are associated with chondrodysplasia,
CC an autosomal recessive disease characterized by skeletal defects
CC caused by abnormalities in the cartilage of limbs, ribs, mandibles
CC and trachea.
CC -i- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -i- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.
CC -i- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D38162; BAA07367.1; -.
CC EMBL; S74574; AAB33439.1; -.
CC PIR; A55648; A55648.
CC MGI; MGI:88446; Col1a1.
CC InterPro; IPR008161; C1g_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR008985; ConA_like lec.gl.
CC InterPro; IPR000885; Fib.Collagen_C.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR003129; TSPN.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 16.
CC Pfam; PF02210; TSPN; 1.
CC ProDom; PD000007; C1g_helix; 4.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFI; 1.
CC SMART; SM00282; LamG; 1.
CC SMART; SM00210; TSPN; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Signal; Alternative splicing;
CC Disease mutation.
CC SIGNAL 1 35 POTENTIAL.
FT PROPEP 36 511 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
FT CHAIN 512 1561 COLLAGEN ALPHA 1(XI) CHAIN.
FT PROPEP 1562 1804 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 37 228 TSP N-TERMINAL.
FT DOMAIN 229 417 NONHELICAL REGION.
FT DOMAIN 418 506 TRIPLE-HELICAL REGION (INTERRUPTED).
FT DOMAIN 507 509 SHORT NONHELICAL SEGMENT.
FT DOMAIN 510 527 TELOPEPTIDE.
FT DOMAIN 528 1540 TRIPLE-HELICAL REGION.
FT DOMAIN 1541 1561 NONHELICAL REGION (C-TERMINAL).
FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 610 610 CROSSLINKING.
FT SITE 1450 1450 CROSSLINKING.
FT VARSPIC 329 413 Missing (in isoform Short).
FT VARIANT 189 195 LDRSES -> SIEVDR (IN CHONDRODYSPLASIA).
FT VARIANT 196 1804 MISSING (IN CHONDRODYSPLASIA).
SQ SEQUENCE 1804 AA; 180963 MW; FE2DB9D8D1E4219A CRC64;

Query Match      87.6%; Score 945; DB 1; Length 1804;
Best Local Similarity 87.6%; Pred. No. 1.8e-47;
Matches 163; Conservative 4; Mismatches 19; Indels 0; Gaps 0;

Qy 1 GIRLKGKGEKGDGPGFGKDMGIKDRGIGPPGPRGDEGPGKGGNGDGPPL 60
|-:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

Db 767 GVRGLKSGKEGDEGPPGFKDMLKXDRGEVCQVPRGEDGPEGPKGRAGTGDGPPS 826
Qy 61 GPPGEKGLGVPLPGYPGQPKSGISGFPGPGANGKGGRTGPKGPRGQRGPTGPR 120
Db 827 GQAGEKGLGVPLGYPGQPKSGTGFPGFGANGKGGRTGPKGPRGQRGPTGPR 886
Qy 121 GERGPRGITGKPGKGNSSGDPAGPGRGPNQPGQPTGPPGKPPGPKDGLPGHP 180
Db 887 GSRGARPTGKPGPKTSGGDPGPPGPRGQPGQPVGPGPKPPGAGKDLGPH 946
Qy 181 GQRGET 186
Db 947 GQRGET 952
RESULT 5
CA2B HUMAN
ID CA2B HUMAN STANDARD; PRT: 1736 AA.
AC P13942; Q07751; Q13271; Q13272; Q13273; Q39866; Q9UIP9;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN COL11A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032717; PubMed=7559422;
RA Vuorio M.M., Pihlajamaa T., Vandenbergh P., Prockop D.J.,
Ala-Kokko L.;
RT "The human COL11A2 gene structure indicates that the gene has not
evolved with the genes for the major fibrillar collagens";
J. Biol. Chem. 270:22873-22881(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Tubby B.;
RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 59-807 FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93314796; PubMed=8325374;
RA Zhidkova N.I., Brewton R.G., Wayne R.;
RT "Molecular cloning of PARP (proline/arginine-rich protein) from human
cartilage and subsequent demonstration that PARP is a fragment of the
NH2-terminal domain of the collagen alpha 2(XI) chain";
FEBS Lett. 326:25-28(1993).
RN [4]
RP SEQUENCE OF 730-1690 FROM N.A.
RX MEDLINE=96435918; PubMed=2760050;
RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,
van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;
RT "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of
cDNA and genomic DNA reveals characteristics of a fibrillar collagen
with differences in genomic organization";
J. Biol. Chem. 264:13910-13916(1989).
RN [5]
RP SEQUENCE OF 1-537 FROM N.A.
RX MEDLINE=96435918; PubMed=8838804;
RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;
RT "The human alpha 2(XI) collagen gene (COL11A2): completion of coding
information, identification of the promoter sequence, and precise
localization within the major histocompatibility complex reveal
overlap with the KES5 gene";
Genomics 32:401-412(1996).
RN [6]
RP ALTERNATIVE SPLICING.
RX MEDLINE=95238468; PubMed=7721876;
RA Zhidkova N.I., Justice S.K., Wayne R.;
RT "Alternative mRNA processing occurs in the variable region of the
pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";

RL J. Biol. Chem. 270:9486-9493(1995).
RN [7]
RP DISEASE.
RX MEDLINE=20143361; PubMed=10677296;
RA Melkonian M., Brunner H.G., Manouvrier S., Hennekam R.,
Superti-Furga A., Kaeerlaeinen H., Pauli R.M., van Esen T.,
Warman M.L., Bonaventura J., Miny P., Ala-Kokko L.;
RT "Autosomal recessive disorder otospondylocomephalodysplasia is
associated with loss-of-function mutations in the COL11A2 gene";
Am. J. Hum. Genet. 66:368-377(2000).
RN [8]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
associated collagen (type IX), and network-forming collagen (type X)
cause a spectrum of diseases of bone, cartilage, and blood vessels";
Hum. Mutat. 9:300-315(1997).
RN [9]
RP VARIANT OSMED ARG-661.
RX MEDLINE=95163096; PubMed=7859284;
RA Vikkula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,
Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,
van den Hoogen F.H.J., Kopers H.-H., Wayne R., Cheah K.S.E.,
Olsen B.R., Warman M.L., Brunner H.G.;
RT "Autosomal dominant and recessive osteochondrodysplasias associated
with the COL11A2 locus";
Cell 80:431-437(1995).
RN [10]
RP VARIANTS GLY-593; LYS-924; LEU-879; THR-1316 AND GLN-1600.
RX MEDLINE=98254467; PubMed=9585596;
RA Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,
Yone K., Matsunaga S., Otterud B., Inoue I., Leppert M.;
RT "Genetic mapping of ossification of the posterior longitudinal
ligament of the spine";
Am. J. Hum. Genet. 62:1460-1467(1998).
RN [11]
RP VARIANT WZS GLU-955.
RX MEDLINE=99021942; PubMed=9805126;
RA Pihlajamaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,
Giedion A., Wiesbauer P., Spranger J., Ala-Kokko L.;
RT "Heterozygous glycine substitution in the COL11A2 gene in the original
patient with the Weissenbacher-Zweymueller syndrome demonstrates its
identity with heterozygous OSMED (nonocular Stickler syndrome)";
Am. J. Med. Genet. 80:115-120(1998).
RN [12]
RP VARIANT STL3 940-GLY--PRO-948 DEL.
RX MEDLINE=98165506; PubMed=9506662;
RA Sirko-Osada D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,
Robin N.H.;
RT "Stickler syndrome without eye involvement is caused by mutations in
COL11A2, the gene encoding the alpha-2(XI) chain of type XI
collagen";
J. Pediatr. 132:368-371(1998).
RN [13]
RP VARIANTS DFNA13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.
RX MEDLINE=20047768; PubMed=10581026;
RA McGuire W.T., Prasad S.D., Griffith A.J., Kunst H.P.M., Green G.E.,
Shargel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E.,
King M.-C., Brunner H.G., Cremers C.W.R.J., Takasus M., Li S.-W.,
Arita M., Wayne R., Prockop D.J., Van Camp G., Smith R.J.H.;
RT "Mutations in COL11A2 cause non-syndromic hearing loss (DFNA13)";
Nat. Genet. 23:413-419(1999).
CC -!- FUNCTION: May play an important role in fibrillogenesis by
controlling lateral growth of collagen II fibrils.
CC -!- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),
alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational
modification of alpha 1(XI). Alpha 1(XI) can also be found instead
of alpha 3(XI)=1(XI).
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=8;
Comment=Isoforms lack exons 6, 7 or 8 or a combination of these
exons. Experimental confirmation may be lacking for some

CC isoforms;
 CC Name=1; IsoId=P13942-1; Sequence=Displayed;
 CC Name=2; IsoId=P13942-2; Sequence=VSP_001167;
 CC Name=3; IsoId=P13942-3; Sequence=VSP_001168;
 CC Name=4; IsoId=P13942-4; Sequence=VSP_001169;
 CC Name=5; IsoId=P13942-5; Sequence=VSP_001167, VSP_001168;
 CC Name=6; IsoId=P13942-6; Sequence=VSP_001167, VSP_001169;
 CC Name=7; IsoId=P13942-7; Sequence=VSP_001168, VSP_001169;
 CC Name=8; IsoId=P13942-8; Sequence=VSP_001167, VSP_001168;
 CC -1- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- PTM: A disulfide-bonded peptide called proline/arginine-rich protein or PARP is released from the amino terminus during extracellular processing and is subsequently retained in the cartilage matrix from which it can be isolated in significant amounts.
 CC -1- DISEASE: Defects in COL11A2 are the cause of Stickler syndrome type 3 (STL3) [MIM:184840]. STL3 is an autosomal dominant disorder characterized by oro-facial, auditory and skeletal manifestations, such as midfacial hypoplasia, cleft palate, osteoarthritis, and sensorineural hearing loss. Differently from Stickler syndrome type 1 and 2, no ocular involvement is observed. This disorder is also referred to as Stickler-like syndrome or non-ocular Stickler syndrome.
 CC -1- DISEASE: Defects in COL11A2 are the cause of autosomal recessive otospondylocheophaeal dysplasia (OSMED) [MIM:215150], a skeletal dysplasia accompanied by severe hearing loss. The phenotype overlaps that of autosomal dominant skeletal disorders (Stickler and Marshall syndromes) but can be distinguished by disproportionately short limbs and lack of ocular involvement.
 CC -1- DISEASE: Defects in COL11A2 are the cause of Weissenbacher-Zweymueller syndrome (WZS) [MIM:277610], an autosomal dominant disorder allelic with STL3 and OSMED. WZS is also referred to as heterozygous OSMED.
 CC -1- DISEASE: Defects in COL11A2 are the cause of autosomal dominant nonsyndromic sensorineural deafness type 13 (DFNA13) [MIM:601868]. Affected individuals experience progressive hearing loss beginning in the second to fourth decades, eventually making use of amplification mandatory.
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -1- DATABASE: NAME=Hereditary hearing loss homepage; NOTE=Gene page; WWW="http://www.uita.ac.be/dnalab/hhh/".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32169; AAC50214.1; -
 CC EMBL; U32169; AAC50213.1; -
 CC EMBL; U32169; AAC50215.1; -
 CC EMBL; AL031228; CAA20240.1; -
 CC EMBL; L18987; AAC35498.1; -
 CC EMBL; J04974; AAC20203.1; -
 CC EMBL; U41069; AAC17464.1; -
 CC EMBL; U41065; AAC17464.1; JOINED.
 CC EMBL; U41065; AAC17464.1; JOINED.
 CC EMBL; U41067; AAC17464.1; JOINED.
 CC Genew; HGNC:2187; COL11A2.

Query Match

84.1%; Score 907; DB 1; Length 1736;

Best Local Similarity 85.4%; Pred. No. 2.6e-45;
 Matches 158; Conservative 7; Mismatches 20; Indels 0; Gaps 0;
 QY 1 GIRGLKGTGKGEKGDGPGFKGDMGIKGRGEIGPPGPRGEDGEGPKRGSGNGDPGPL 60
 DB 727 GIRGLKGHKGEKGDGPGFKGDIKVGDRGEVGVPSRGEDGEGPKRGRTGTGDPGP 786
 QY 61 GPPGEXKGLGVPLPGYPORQPGKSGIGPPGPPGANGKGGRTTPKPPRGORGPTGPR 120
 DB 787 GLMGEXKGLGVPLPGYPORQPGKSLGFGPPGASGKRGARLSGSKSGPRGRTGPR 846
 QY 121 GERGPRGITGKPGKNSGDDGPPGPRGPNPGPOGTGFGPKGPPGPGKDGGLPGHP 180
 DB 847 GQGRPRGATGKSGAKTSGDGHGPPGPRGKLPQPGNGFGPGPPGPGKDGGLPGHP 906
 QY 181 GORGE 185
 DB 907 GORGE 911
 RESULT 6
 CA2B MOUSE
 ID CA2B MOUSE STANDARD; PRT; 1736 AA.
 AC Q64739; O61432; Q521W0;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 2(XI) chain precursor.
 GN COL11A2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 7).
 RC STRAIN=129/SvJ.
 RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L., Hall J., Lasky S., Hood L.;
 RA "Sequence of the mouse major histocompatibility locus class II region."
 RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-1678 FROM N.A. (ISOFORM 7).
 RC STRAIN=EVB/N, and 129/Sv; TISSUE=Cartilage;
 RX MEDLINE=97133795; PubMed=8981332;
 RA Vandenberg P., Vuoristo M.M., Ala-Kokko L., Prockop D.J.;
 RT "The mouse coll1a2 gene. Some transcripts from the adjacent rrx-beta gene extend into the coll1a2 gene."
 RL Matrix Biol. 15:359-367(1996).
 RN [3]
 RP SEQUENCE OF 1-624 FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).
 RC STRAIN=129/Sv;
 RX MEDLINE=95138212; PubMed=7836472;
 RA Tsumaki N., Kimura T.;
 RT "Differential expression of an acidic domain in the amino-terminal propeptide of mouse pro-alpha2(XI) collagen by complex alternative splicing."
 RL J. Biol. Chem. 270:2372-2378(1995).
 RN [4]
 RP SEQUENCE OF 1-8 FROM N.A.
 RC STRAIN=129/Sv; TISSUE=Liver;
 RX MEDLINE=96427460; PubMed=8830784;
 RA Tsumaki N., Kimura T., Matsui Y., Ochi T.;
 RT "Separable cis-regulatory elements that contribute to tissue- and site-specific alpha 2(XI) collagen gene expression in the embryonic mouse cartilage."
 RL J. Cell Biol. 134:1573-1582(1996).
 CC -1- FUNCTION: May play an important role in fibrillogenesis by controlling lateral growth of collagen II fibrils (By similarity).
 CC -1- SUBUNIT: Trimers composed of three different chains: alpha 1(XI), alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational modification of alpha 1(II). Alpha 1(IV) can also be found instead of alpha 3(XI)=1(II) (By similarity).
 CC


```
CC CC -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 1(XI) CHAINS.
CC CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AF177941; AAF59902.1; -.
CC CC PIR; S20375; S20375.
CC CC Genew; HGNC:14864; COL5A3.
CC CC MM; 120216; -.
CC CC GO; GO:0005588; C:collagen type V; TAS.
CC CC GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
CC CC DR InterPro; IPR008161; Clg helix.
CC CC DR InterPro; IPR008160; Collagen.
CC CC DR InterPro; IPR000885; Fib_collagen_C.
CC CC DR InterPro; IPR003129; TSPN.
CC CC Pfam; PF01410; COLFI; 1.
CC CC Pfam; PF01391; Collagen; 17.
CC CC Pfam; PF02210; TSPN; 1.
CC CC ProDom; PD000007; Clg helix; 2. C; 1.
CC CC ProDom; PD002076; Fib_collagen_C; 1.
CC CC SMART; SM00038; COLFI; 1.
CC CC SMART; SM00210; TSPN; 1.
CC CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC CC Collagen; Signal.
CC CC SIGNAL 1 29 POTENTIAL.
CC CC CHAIN 30 1745 COLLAGEN ALPHA 3(V) CHAIN.
CC CC DOMAIN 31 210 TSP N-TERMINAL.
CC CC DOMAIN 211 391 NONHELICAL REGION.
CC CC DOMAIN 392 1489 TRIPLE-HELICAL REGION.
CC CC DOMAIN 1490 1745 NONHELICAL REGION.
CC CC SEQUENCE 1745 AA; 172051 MW; 5E8FF97135397AC1 CRC64;
CC CC
CC CC Query Match 68.3%; Score 737; DB 1; Length 1745;
CC CC Best Local Similarity 72.4%; Pred. No. 1.3e-35;
CC CC Matches 134; Conservative 12; Mismatches 39; Indels 0; Gaps 0;
CC CC
QY 1 GIRLKTKEKEDGPPGKGMKIGDRGETGPPGPRGDEGPEGPKGGNGDPGL 60
Db 719 GNRGLQGEKEDGPPGKGMKIGDRGETGPPGPRGDEGPEGPKGGNGDPGL 778
QY 61 GPPEKGLVGLPGVPGKQKSGISGFFPGKANGKGGRTPGKPGPRGPGTGP 120
Db 779 GSAGEKGLVGLPGVPGKQKSGISGFFPGKANGKGGRTPGKPGPRGPGTGP 838
QY 121 GERGPRGIGTKPGKNGSGDGPAGPPGERGPNPGQPTGPPGKPPGKGLPGHP 180
Db 839 GERGPRGIGTKPGKNGSGDGPAGPPGERGPNPGQPTGPPGKPPGKGLPGHP 898
QY 181 GQGE 185
Db 899 GQGE 903
CC CC
CC CC RESULT 8
CC CC CA13_BOVIN STANDARD; PRT; 1049 AA.
CC CC ID CA13_BOVIN
CC CC AC P04258;
CC CC DT 20-MAR-1987 (Rel. 04, Created)
CC CC DT 20-MAR-1987 (Rel. 04, Last sequence update)
CC CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC CC DE Collagen alpha 1(III) chain.
CC CC GN COL3A1.
CC CC OS Bos taurus (Bovine).
CC CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC CC OC Bovidae; Bovinae; Bos.
CC CC NCBI_TaxID=9913;
```

```
RN RP SEQUENCE OF 1-242.
RX MEDLINE=80026026; PubMed=488906;
RA Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wachter E.,
RA Kuhn K.;
RT "The covalent structure of calf skin type III collagen. I. The amino
RT acid sequence of the amino terminal region of the alpha 1(III) chain
RT (positions 1-222).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
RN RP [2]
RX MEDLINE=80026027; PubMed=488907;
RA Dawes H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. II. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CBI.8,10,2
RT (positions 223-402).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
RN RP [3]
RX MEDLINE=80026028; PubMed=488908;
RA Bentz H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. III. The
RT amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT (positions 403-551).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
RN RP [4]
RX MEDLINE=80026029; PubMed=488909;
RA Lang H., Glangville R.W., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. IV. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT (positions 552-788).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
RN RP [5]
RX MEDLINE=80026030; PubMed=488910;
RA Dawes H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. V. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT (position 789-927).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
RN RP [6]
RX MEDLINE=80026031; PubMed=488911;
RA Allmann H., Fietzek P.P., Glangville R.W., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. VI. The amino
RT acid sequence of the carboxyterminal cyanogen bromide peptide alpha
RT 1(III)CB9B (positions 928-1028).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
CC CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
CC CC along with type I collagen.
CC CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC CC linked to each other by interchain disulfide bonds. Trimers are
CC CC also cross-linked via hydroxylsines.
CC CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC CC PIR; A02862; CGB078.
CC CC InterPro; IPR008161; Clg helix.
CC CC InterPro; IPR008160; Collagen.
CC CC InterPro; IPR001007; VWF-C.
CC CC Pfam; PF01391; Collagen; 17.
CC CC ProDom; PD000007; Clg helix; 3.
CC CC PROSITE; PS01208; VWF_C_1; PARTIAL.
CC CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC CC Glycoprotein; Collagen.
CC CC DOMAIN 1 14 NONHELICAL REGION (N-TERMINAL).
CC CC DOMAIN 15 1040 TRIPLE-HELICAL REGION.
CC CC DOMAIN 1041 1049 NONHELICAL REGION (C-TERMINAL).
CC CC MOD_RES 95 95 HYDROXYLATION.
CC CC MOD_RES 107 107 HYDROXYLATION.
CC CC MOD_RES 119 119 HYDROXYLATION.
CC CC MOD_RES 938 938 HYDROXYLATION.
CC CC MOD_RES 950 950 HYDROXYLATION.
```

```
FT CARBOHYD 107 107 O-LINKED (GAL. . .).
FT CARBOHYD 950 950 O-LINKED (GAL. . .).
FT DISULFID 1040 1040 INTERCHAIN.
FT DISULFID 1041 1041 INTERCHAIN.
SQ SEQUENCE 1049 AA, 93651 MW, 8EBC3D1C66EC9A3 CRC64;

Query Match 47.2%; Score 509.5; DB 1; Length 1049;
Best local similarity 38.5%; Pred. No. 8.8e-23;
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;

QY 1 GIRLXGT-----KGEKGEDGFGFKDGMGKIDRGEI---GPPGGRGED 42
Db 483 GLQLPOTSGPPGNGKPGEPGPKGEAGAPGIPGGKGDSCAPGERGPPGAGGPPGPRGA 542
QY 43 GPEGPKG-----RGGN----- 54
Db 543 GPGPEGGAAGPPGPPGSGAGTGLQGMFGERGGPGPKDKGEFGSSGVDGAPGKD 602
QY 55 -----GDPGLGPPGEGKGLGVPLPGYP-----GRQPKGSIQPPGPPGAN 96
Db 603 GPRGPTGPIGPPGAGQPGDKGESGAPGVFGIAGPRGGPGERGEQGPFGPAGFPAGQN 662
QY 97 GEKGGRTGKPKGPRCQ-----RGPTGPRGERGPRGITKPGPKGNSGSD 141
Db 663 GEPGKGERGAPGERGEGGPPGAAGPAGSGPAGPPGQGVKGERSGCGGAAGFPGR 722
QY 142 GPAGPPGERGPNPGQPTGPPGPKGPPGPPGKDLGHPGQRG 184
Db 723 GPPGPGSGNPGPPGSSGAPGDKGPPGPGSGNAPGSGISG 765

RESULT 9
CAS4_HUMAN STANDARD; PRT; 1685 AA.
ID CA54_HUMAN
AC P29470: Q16006; Q16126;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5(IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
[2]
SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RP TISSUE=Kidney;
RC MEDLINE=92316923; PubMed=1352287;
RX Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RA "Complete amino acid sequence of the human alpha 5 (IV) collagen
chain and identification of a single-base mutation in exon 23
RT converting glycine 521 in the collagenous domain to cysteine in an
RT Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
[3]
SEQUENCE OF 85-1685 FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjola-Erkkonen E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the
RT carboxyl-terminal domain of a new type IV collagen chain, alpha
RT 5(IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
[4]
SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
```

```
RT "Characterization of the 3' half of the human type IV collagen alpha
5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
[5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeslytye M., Shows T.B.,
RA Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with
restricted kidney distribution and assignment of its gene to the
RT locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
[6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjola-Erkkonen E.R., Kadri A.S., Goddard A.D.,
RA Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
RT to the region of the X chromosome containing the Alport syndrome
RL locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
[7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,
RA Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
[8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
RA Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood
RT cells: a complex mutation in the COL4A5 gene of an Alport patient
RT deletes the Nc1 domain.";
RL Kidney Int. 44:1316-1321(1993).
[9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
[10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
RA Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
RT conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
[11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,
RA Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5
RT (IV) chain associated with X-linked Alport syndrome: characterization
RT of the mutation by direct sequencing of PCR-amplified lymphoblast
RT cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
[12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
RA Rizzoni G.P., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to
RT glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
[13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
RA Tryggvason K., Haggema-Schouten W.A.G., Roodvoets A.P., Rascher W.,
RA van Oost B.A., Smeets H.J.N.;
```

RT "Identification of four novel mutations in the COL4A5 gene of
 RT patients with Alport syndrome."
 RL Genomics 17:485-489(1993).
 RN [14]
 RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796;
 RP ARG-869; ARG-872 AND CYS-1241.
 RX MEDLINE=95322976; PubMed=7595631;
 RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
 RA "Detection of 12 novel mutations in the collagenous domain of the
 RT COL4A5 gene in Alport syndrome patients."
 RL Hum. Mutat. 5:197-204(1995).
 RN [15]
 RP VARIANT AS ARG-1649.
 RX MEDLINE=96213750; PubMed=8651292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 RA Denison J.C., Fain P.R., Gregory M.C.;
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 RT common in the western United States."
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [16]
 RP VARIANTS AS.
 RX MEDLINE=96213754; PubMed=8651296;
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
 RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
 RA Scolari F., Sesca A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
 RA Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 RT exons of the COL4A5 gene."
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [17]
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
 RP MET-1428.
 RX MEDLINE=97094179; PubMed=8940267;
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
 RA Glatas I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
 RA Gubler M.-C., Antignac C.;
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
 RT syndrome."
 RL Am. J. Hum. Genet. 59:1221-1232(1996).
 RN [18]
 RP VARIANT AS ASP-1498.
 RX MEDLINE=96233932; PubMed=8829632;
 RA Tverskaya S., Bobrylnina V., Tsalykova F., Ignatova M.,
 RA Krasnopolskaya X., Evgrafov O.;
 RT "Substitution of A1498D in noncollagen domain of $\alpha 5(\text{IV})$ collagen
 RT chain associated with adult-onset X-linked Alport syndrome."
 RL Hum. Mutat. 7:149-150(1996).
 RN [19]
 RP VARIANT AS GLN-1677.
 RX MEDLINE=97295089; PubMed=9150741;
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 RT syndrome and COL4A5 R1677Q."
 RL Hum. Genet. 99:681-684(1997).
 RN [20]
 RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
 RP AND ASP-1596.
 RX MEDLINE=98112435; PubMed=9452056;
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
 RA Trivelli A., Pinciaroli A.R., Regaiolo M., Rizzoni G.F., de Marchi M.;
 RT "Missense mutations in the COL4A5 gene in patients with X-linked
 RT Alport syndrome."
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).
 RN [21]
 RP VARIANTS AS VAL-420; 456-PRO--PRO-458 DEL; ASP-573; ASP-624; ASP-635;
 RP 802-GLY--PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;
 RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
 RX MEDLINE=99063529; PubMed=9848783;
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
 RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,
 RA Springate J., Shows T.B., Petersson E., Tryggvason K.;
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected

RT Alport syndrome using PCR and direct DNA sequencing."
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
 RN [22]
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
 RX MEDLINE=20030197; PubMed=10561141;
 RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
 RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
 RT "Detection of mutations in the COL4A5 gene in over 90% of male
 RT patients with X-linked Alport's syndrome by RT-PCR and direct
 RT sequencing."
 RL Am. J. Kidney Dis. 34:854-862(1999).
 RN [23]
 RP VARIANT AS ARG-822.
 Query Match 46.5%; Score 501.5; DB 1; Length 1685;
 Best Local Similarity 47.2%; Pred. No. 3.7e-22;
 Matches 108; Conservative 13; Mismatches 63; Indels 45; Gaps 7;
 QY 1 GTRGLKGTGKEGDEDFPGKGMGIKGRGE-----TGPDPREDGP--EGPKG 49
 DB 1193 GPPGLFGLSGQKGDGLPGIPGNPLGPKGPFHGFPGVQVQPPGSPGPALEGPKG 1252
 QY 50 RCGPND-----PGPLGPP-----GEKGLGVPLGVY---GROGPKSGIFPG 91
 DB 1253 NFGPQGPGRGGLPGEPPGLPGNGIKGEKNPQGLGVLGKDGQGPGLQGNPG 1312
 QY 92 PFGANGEKGRGTGPKGPRGQRGPTGPRGERGPR---GITGKPGP-----K 135
 DB 1313 RPLGLNMGKGDGLPGVPGFPGMKPSGVPSAGPEGEGLIGPPGLPGSPGQSIIK 1372
 QY 136 GNSGDDGPPAGPPGERGPNQPGTTPPGKGPFGPPGKGLPCHHPQQR 184
 DB 1373 GDAGPPGIPQPLGKLPQGPQGLPPTGTPGTPGDPGRNGLPGLFDGAGG 1421
 RESULT 10
 CAL4_MOUSE
 ID CA14_MOUSE STANDARD; PRT; 1669 AA.
 AC P02463;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCHI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89197932; PubMed=2703490;
 RA Muthukumar G., Blumberg B., Kurkinen M.;
 RT "The complete primary structure for the alpha 1-chain of mouse
 RT collagen IV. Differential evolution of collagen IV domains."
 RL J. Biol. Chem. 264:6310-6317(1989).
 RN [2]
 RP SEQUENCE OF 1-1154 FROM N.A.
 RX MEDLINE=88112221; PubMed=3338568;
 RA Wood L., Theriault N., Vogeli G.;
 RT "CDNA clones completing the nucleotide and derived amino acid
 RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
 RT from mouse."
 RL FEBS Lett. 227:5-8(1988).
 RN [3]
 RP SEQUENCE OF 1149-1424 FROM N.A.
 RX MEDLINE=86301886; PubMed=3755692;
 RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
 RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
 RT synthetic oligodeoxynucleotide."
 RL Gene 43:301-304(1986).
 RN [4]
 RP SEQUENCE OF 1276-1669 FROM N.A.

RX MEDLINE=85127033; PubMed=2578961;
RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224 (1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499 (1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657 (1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277 (1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Buelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682 (1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Buelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709 (1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03758; AAA37439.1; -
DR EMBL; M23333; AAA51625.1; -
DR EMBL; J04694; AAA50292.1; -
DR EMBL; X06777; CAA29946.1; -

DR EMBL; X02201; CAA26132.1; -
DR EMBL; M15832; AAA37340.1; -
DR EMBL; M14042; AAA37342.1; -
DR EMBL; M12879; AAA37343.1; -
DR EMBL; M13024; -; NOT ANNOTATED CDS.
DR EMBL; M13025; -; NOT ANNOTATED CDS.
DR EMBL; M13026; AAA37344.1; -
DR EMBL; M13027; AAA37345.1; -
DR EMBL; M13043; AAA37346.1; -
DR EMBL; J04448; AAA37437.1; -
DR PIR; A33525; CGMS4B.
DR MGD; MGI:88454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1q helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; C1q helix; 6.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1565 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
FT DISULFID 1616 1822 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 26 26 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 Q -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 Q -> H (IN REF. 2).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 46.4%; Score 501; DB 1; Length 1669;
Best Local Similarity 46.2%; Pred. No. 3.9e-22;
Matches 108; Conservative 14; Mismatches 64; Indels 48; Gaps 7;

QY 1 GIRLKGTKGKGBDGFPGFKDNGI---KGDREIGPPGRGBDG-----PEGPKG 49
DB 1177 GFPGSKGDKSGEVEGFGFGLAGSPGIPGVKGEQGMGPPGQPGPLGTGHPVGEKPG 1236
QY 50 RGGPNGD-----PGLGPP-----GSKGLGVPLGVPGRGQPKSGIGFPGF-- 92
DB 1237 DRGQGGQGLPGHPGMPGPPGPGINGPKDKNQGWPGAFVPGPKDQFGQMPGIGG 1296
QY 93 -PGANGKGRGTGKPGPRGQRGTGPRGERGRTGTGKPGKNGSGDGPGP----- 146
DB 1297 SFGITGSGKDMGLPGVPGFQCGKGLPGLGVKQDQGVPGKLGQPGPPGPGYDVK 1356
QY 147 --PBERGNGQGRTGPPGPKGP-----PGPKGDKGLPGHPGQRGET 186
DB 1357 GEPLGPGEGPPGLKGLGPPGPGQGVTSVGLPGPGVPGPDGAPGQKGET 1410

RESULT 11
CALA_CHICK
ID CAIA_CHICK
AC P08125; STANDARD; PRT; 674 AA.

assembly and produce Ehlers-Danlos syndrome type I.";
 Hum. Mol. Genet. 7:249-255(1998).
 [8]
 RP VARIANT EDS-II ARG-960.
 RX MEDLINE=98455031; PubMed=9783710;
 RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
 Burrows N.P.;
 RA "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
 RT I.";
 RL J. Med. Genet. 35:846-848(1998).
 CC -!- FUNCTION: Type V collagen is a member of group I collagen
 (fibrillar forming collagen). It is a minor connective tissue
 component of nearly ubiquitous distribution. Type V collagen binds
 to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
 CC -!- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
 most tissues and trimers of one alpha 1(V), one alpha 2(V), and
 one alpha 3(V) chains in placenta.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
 type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
 gravis. EDS-I is an autosomal dominant connective-tissue disorder
 characterized by loose-jointedness and fragile, velvety,
 stretchable, bruiseable skin that heals with peculiar 'cigarette-
 paper' scars. Inheritance is autosomal dominant.
 CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
 type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
 syndrome mitis. Inheritance is autosomal dominant.
 CC -!- SIMILARITY: Contains 1 VWFC domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; J04478; AAA51859.1; -;
 DR ENBL; X04758; CAA28454.1; -;
 DR ENBL; M11718; AAA52058.1; -;
 DR PIR; A31427; CGHU2V.
 DR PIR; IA9A; 18-NOV-98.
 DR Genew; HGNC:2210; COL5A2.
 DR MIM; 120190; -;
 DR MIM; 130000; -;
 DR MIM; 130010; -;
 DR GO; GO:0005588; C:collagen type V; TAS.
 DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR002181; Fibrinogen_C-
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; vwc; 1.
 DR ProDom; PD000007; Clg_helix; 5.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWFC 1; 1.
 DR PROSITE; PS0184; VWFC 2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
 KW Disease mutation; 3D-structure.
 FT SIGNAL 1 26
 FT CHAIN 27 1226 COLLAGEN ALPHA 2(V) CHAIN.
 FT PROPEP 1227 1496 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMATN 39 97 VWFC.
 FT MOD_RES 290 290 HYDROXYLATION.
 FT MOD_RES 293 293 HYDROXYLATION.
 FT MOD_RES 296 296 HYDROXYLATION.

FT	MOD_RES	608	608	HYDROXYLATION.
FT	MOD_RES	614	614	HYDROXYLATION.
FT	VARIANT	960	960	G -> R (in EDS-II). /FTid=VAR_013588.
FT	CONFLICT	292	292	A -> P (IN REF. 6).
FT	CONFLICT	1418	1418	K -> T (IN REF. 3).
FT	CONFLICT	1438	1438	F -> S (IN REF. 3).
FT	CONFLICT	1460	1460	E -> Q (IN REF. 4).
FT	CONFLICT	1496	1496	V -> A (IN REF. 4).
SQ	SEQUENCE	1496 AA;	144720 MW;	82827C17A8644F5A CRC64;

Query Match 45.5%; Score 491; DB 1; Length 1496;
 Best Local Similarity 51.4%; Pred. No. 1.3e-21;
 Matches 95; Conservative 12; Mismatches 78; Indels 0; Gaps 0;

QY	1	GIRGLKGTGKGEKGDGFGPKGDMGKIGDRGTGPPGCGEDGPEGPKGSGGNGDGPGL	60
DB	291	GAPGLPGLKGRHKGLEGPKEVGAPGSGKEAGTGPWGAMGLPGRGNPGERGLGPQ	350
QY	61	GPPEKGLGVPLPGYPRQPKSGIGFPFGANGKGGRTGPKGPRGQGTGPR	120
DB	351	GAPGQGAHGMKPGKPGMGLGIPGSSGPPGNPMKGEAGTGAEGPQGGQGTGPP	410
QY	121	GERGPRGITKPGPKGNSGDDGAPGPRGPRGPNPGTGFPGPKGPPGKGLGCHP	180
DB	411	GPVGSPLGCAIGTGTGTPGKPTGSGTSGPPGSGAGPPGSGPQSGTGPQNSGLP	470
QY	181	GORGE	185
DB	471	GPKGE	475

RESULT 14
 CA16 CHICK
 ID CA16 CHICK STANDARD; PRT; 1019 AA.
 AC P20785;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(VI) chain precursor.
 GN COL6A1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92241293; PubMed=1572359;
 RA Walchli C., Koller E., Trueb J., Trueb B.;
 RT "Structural comparison of the chicken genes for alpha 1(VI) and alpha
 RT 2(VI) collagen.";
 RL Eur. J. Biochem. 205:583-589(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89174602; PubMed=2784434;
 RA Bonaldo P., Russo V., Buccioti F., Bressan G.M., Colombatti A.;
 RT "Alpha 1 chain of chick type VI collagen. The complete cDNA sequence
 RT reveals a hybrid molecule made of one short collagen and three von
 RT Willebrand factor type A-like domains.";
 RL J. Biol. Chem. 264:5575-5580(1989).
 RN [3]
 RP SEQUENCE OF 1-75 FROM N.A.
 RX MEDLINE=93011107; PubMed=1396681;
 RA Koller E., Trueb B.;
 RT "Characterization of the chicken alpha 1(VI) collagen promoter.";
 RL Eur. J. Biochem. 208:769-774(1992).
 CC -!- FUNCTION: Collagen VI acts as a cell-binding protein.
 CC -!- SUBUNIT: Trimers composed of three different chains: alpha 1(VI),
 CC alpha 2(VI), and alpha 3(VI).
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- SIMILARITY: Contains 3 VWFA domains.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:05:48 ; Search time 20.8274 Seconds
(without alignments)
2817.753 Million cell updates/sec

Title: US-09-775-964-6
Perfect score: 1079
Sequence: 1 GRLGLKGTGKGGDFPGF.....PPGPPKGLPGHQRGET 186

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTEMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1079	100.0	1838	4 Q15094	Q15094 homo sapien
2	1071	99.3	1838	11 Q88207	Q88207 mus musculus
3	1071	99.3	1840	11 Q60467	Q60467 cricetus
4	1071	99.3	1840	11 Q9J103	Q9J103 rattus norv
5	1037	96.1	1835	13 Q9IAU4	Q9IAU4 gallus gall
6	945	87.6	1804	11 Q8QWR4	Q8QWR4 mus musculus
7	885	82.0	1820	13 Q91907	Q91907 pagrus majo
8	880	81.6	1827	13 Q8UUM5	Q8UUM5 oryzias lat
9	741	68.7	1053	11 Q920S4	Q920S4 mus musculus
10	741	68.7	1739	11 Q9JLI2	Q9JLI2 mus musculus
11	723	67.0	1737	11 Q9JIO4	Q9JIO4 rattus norv
12	533.5	49.4	1349	13 Q8AW17	Q8AW17 brachydanio
13	509.5	47.2	890	5 Q77087	Q77087 alvinella p
14	506.5	46.9	1860	4 Q81ZC6	Q81ZC6 homo sapien
15	501.5	46.5	886	4 Q9NUB7	Q9NUB7 homo sapien
16	497.5	46.1	888	13 Q90796	Q90796 gallus gall

17	494	45.8	854	4 Q8IVT9	Q8IVT9 homo sapien
18	493.5	45.7	1669	11 Q9QZS0	Q9QZS0 mus musculus
19	493	45.7	1347	4 Q96QB3	Q96QB3 homo sapien
20	492	45.6	642	4 Q9BXR9	Q9BXR9 homo sapien
21	492	45.6	654	4 Q9BXS0	Q9BXS0 homo sapien
22	488.5	45.3	1491	13 Q91718	Q91718 xenopus lae
23	487.5	45.2	998	11 Q8CFM4	Q8CFM4 mus musculus
24	487.5	45.2	1222	11 Q8K173	Q8K173 mus musculus
25	487.5	45.2	1464	11 Q8BK72	Q8BK72 mus musculus
26	487.5	45.2	1464	11 Q8BLW4	Q8BLW4 mus musculus
27	487.5	45.2	1464	11 Q7TT32	Q7TT32 mus musculus
28	487	45.1	675	6 Q9N178	Q9N178 sus scrofa
29	486	45.0	1821	4 Q9H4R9	Q9H4R9 homo sapien
30	485.5	45.0	1491	13 Q7ZTM4	Q7ZTM4 xenopus lae
31	485	44.9	1075	4 Q86X41	Q86X41 homo sapien
32	484.5	44.9	891	11 Q80VS8	Q80VS8 mus musculus
33	484.5	44.9	1486	13 Q91717	Q91717 xenopus lae
34	484	44.9	284	4 Q9UC14	Q9UC14 homo sapien
35	484	44.9	349	4 Q99018	Q99018 homo sapien
36	484	44.9	1497	4 Q9NOK9	Q9NOK9 homo sapien
37	484	44.9	1532	4 Q02802	Q02802 homo sapien
38	483	44.8	1269	13 Q7T2Z7	Q7T2Z7 gallus gall
39	483	44.8	1420	13 Q90W37	Q90W37 gallus gall
40	482.5	44.7	1418	13 Q9W7R9	Q9W7R9 cynops pyrr
41	482.5	44.7	1802	5 Q17163	Q17163 brugia mala
42	482	44.7	666	11 Q99MQ5	Q99MQ5 mus musculus
43	482	44.7	1497	4 Q9UMD9	Q9UMD9 homo sapien
44	482	44.7	1691	11 Q9ESQ2	Q9ESQ2 mus musculus
45	481.5	44.6	280	13 Q91014	Q91014 gallus gall

ALIGNMENTS

RESULT 1

Q15094 Q15094 PRELIMINARY; PRT; 1838 AA.

AC Q15094;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pro-alpha-1 type V collagen.
GN COL5A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92105142; PubMed=172213;
RA Greenspan D.S., Cheng W., Hoffman G.G.;
RT "The pro-alpha 1(V) collagen chain. Complete primary structure,
RT distribution of expression, and comparison with the pro-alpha 1(XI)
RT collagen chain.";
RL J. Biol. Chem. 266:24727-24733 (1991).
DR EMBL; M76729; AAA5993.1; -;
DR PIR; S18802; CGHUIV.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA like lec gl.
DR InterPro; IPR000885; Fib.Collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.

```
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1838 AA; 183610 MW; 5078307F6E00F0BA CRC64;

Query Match 100.0%; Score 1079; DB 4; Length 1838;
Best Local Similarity 100.0%; Pred. No. 8e-79;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPGPGANGKGGRTGKPGPRGQGTGPR 120
DB 799 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPGPGANGKGGRTGKPGPRGQGTGPR 858

QY 61 GPPGKGLGVPLPGVPRGQKSGDGAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 120
DB 859 GPPGKGLGVPLPGVPRGQKSGDGAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 918

QY 121 GERGPRGITGKPGKNSGGDGPAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 180
DB 919 GERGPRGITGKPGKNSGGDGPAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 978

QY 181 GORGET 186
DB 979 GORGET 984

RESULT 2
O88207 PRELIMINARY; PRT; 1838 AA.
AC O88207;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen a1(V).
GN COL5A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98250615; PubMed=9582436;
RA Wu Y.-L., Sumiyoshi H., Khaleduzaman M., Ninomiya Y., Yoshioka H.;
RT "cDNA sequence and expression of the mouse alpha1(V) collagen gene.";
RL Biochim. Biophys. Acta 1397:275-284(1998).
DR EMBL; AB009993; BAA28786.1; -
DR MGD; MGI:88457; Col5a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1838 AA; 183691 MW; 7A520B23D18517B3 CRC64;

Query Match 99.3%; Score 1071; DB 11; Length 1838;
Best Local Similarity 99.5%; Pred. No. 3.5e-78;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPGPRGANGKGGRTGKPGPRGQGTGPR 60
DB 799 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPGPRGANGKGGRTGKPGPRGQGTGPR 858
```

```
QY 61 GPPGKGLGVPLPGVPRGQKSGDGAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 120
DB 859 GPPGKGLGVPLPGVPRGQKSGDGAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 918

QY 121 GERGPRGITGKPGKNSGGDGPAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 180
DB 919 GERGPRGITGKPGKNSGGDGPAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 978

QY 181 GORGET 186
DB 979 GORGET 984

RESULT 3
Q60467 PRELIMINARY; PRT; 1840 AA.
AC Q60467;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pro-alpha-1 type V collagen.
OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92105142; PubMed=1722213;
RA Greenspan D.S., Cheng W., Hoffman G.G.;
RT "The pro-alpha-1(V) collagen chain: Complete primary structure,
RT distribution of expression, and comparison with the pro-alpha-1(XI)
RT collagen chain.";
RL J. Biol. Chem. 266:24727-24733(1991).
DR EMBL; M76730; AAA37002.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1840 AA; 184174 MW; 32C56821EF64CE75 CRC64;

Query Match 99.3%; Score 1071; DB 11; Length 1840;
Best Local Similarity 99.5%; Pred. No. 3.6e-78;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPGPRGANGKGGRTGKPGPRGQGTGPR 60
DB 801 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPGPRGANGKGGRTGKPGPRGQGTGPR 860

QY 61 GPPGKGLGVPLPGVPRGQKSGDGAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 120
DB 861 GPPGKGLGVPLPGVPRGQKSGDGAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 920

QY 121 GERGPRGITGKPGKNSGGDGPAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 180
DB 921 GERGPRGITGKPGKNSGGDGPAGPPGPRGPNQPGTGFPGKPGPPGKDGILPGHP 980

QY 181 GORGET 186
```

```
Db          981 GORGET 986
|||||
RESULT 4
Q9J103      PRELIMINARY;      PRT; 1840 AA.
ID Q9J103
AC Q9J103
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 1 type V collagen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley;
RX MEDLINE=20428740; PubMed=10852920;
RA Chernousov M.A., Rothblum K., Tyler W.A., Stahl R.C., Carey D.J.;
RT "Schwann cells synthesize type V collagen that contains a novel alpha
RT 4 chain. molecular cloning, biochemical characterization, and high
RT affinity heparin binding of alpha4(V) collagen.";
RL J. Biol. Chem. 275:28208-28215(2000).
DR ENBL; AF272662; AAF76433.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lect_g1.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000508; Peptidase_S26.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00761; SPASE_I_3; 1.
KW Collagen.
SQ SEQUENCE 1840 AA; 183986 MW; AD38F5FF868B923C CRC64;

Query Match          99.3%; Score 1071; DB 11; Length 1840;
Best Local Similarity 99.5%; Pred. No. 3.6e-78;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGPGFGKDMGKDRGEIGPPGPRGEDGPGKRGSGNGDPGPL 60
Db          801 GIRLKGTKGKEDGPGFGKDMGKDRGEIGPPGPRGEDGPGKRGSGNGDPGPL 860

QY 61 GPPGKGLGVPLPGYPGQPKSGISGFFPGANGKGGRTGCKPGPRGQGTGPR 120
Db          861 GPTGKGLGVPLPGYPGQPKSGISGFFPGANGKGGRTGCKPGPRGQGTGPR 920

QY 121 GERGPGRITGKPGPKNGSGDGPAGPPGERGPNPGQGTGFPKPGPPGPKDGLPGHP 180
Db          921 GERGPGRITGKPGPKNGSGDGPAGPPGERGPNPGQGTGFPKPGPPGPKDGLPGHP 980

QY 181 GORGET 186
|||||
Db          981 GORGET 986
|||||
RESULT 5
```

```
Q9IAU4      PRELIMINARY;      PRT; 1835 AA.
ID Q9IAU4
AC Q9IAU4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 1 (V) collagen.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
NCBI_TaxID=9031;
[1]
SEQUENCE FROM N.A.
RC STRAIN=bred White Leghorn;
RX MEDLINE=20068042; PubMed=10601735;
RA Gordon M.K., Marchant J.K., Foley J.W., Igoe F., Gibney E.P.,
RA Nah H.D., Barembaum M., Myers J.C., Rodriguez E., Dublet B.,
RA van der Rest M., Linsenmayer T.F., Upholt W.B., Birk D.E.;
RT "Complete primary structure of the chicken alpha1(V) collagen chain.";
RL Matrix Biol. 18:481-486(1999).
DR PIR; A55047; A55047.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lect_g1.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 19.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1835 AA; 184234 MW; D05B9D71022D4B2 CRC64;

Query Match          96.1%; Score 1037; DB 13; Length 1835;
Best Local Similarity 96.2%; Pred. No. 2e-75;
Matches 179; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGPGFGKDMGKDRGEIGPPGPRGEDGPGKRGSGNGDPGPL 60
Db          795 GVRGLKGTKGKEDGPGFGKDMGKDRGEIGPPGPRGEDGPGKRGSGNGDPGPL 854

QY 61 GPPGKGLGVPLPGYPGQPKSGISGFFPGANGKGGRTGCKPGPRGQGTGPR 120
Db          855 GPAGKGLGVPLPGYPGQPKSGISGFFPGANGKGGRTGCKPGPRGQGTGPR 914

QY 121 GERGPGRITGKPGPKNGSGDGPAGPPGERGPNPGQGTGFPKPGPPGPKDGLPGHP 180
Db          915 GERGPGRITGKPGPKNGSGDGPAGPPGERGPNPGQGTGFPKPGPPGPKDGLPGHP 974

QY 181 GORGET 186
|||||
Db          975 GORGET 980
|||||
RESULT 6
Q80WR4      PRELIMINARY;      PRT; 1804 AA.
ID Q80WR4
AC Q80WR4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collial protein.
OS Mus musculus (Mouse).
```


OS Oryzias latipes (Medaka fish) (Japanese ricefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HG-RR;
RA Matsuo M.Y., Asakawa S., Shimizu N., Kimura H., Nonaka M.;
RT "Nucleotide Sequence of the MHC Class I Region of a Teleost, the
Medaka";
RL Immunogenetics 0:0-0(2002).
DR EMBL; AB073376; BAB83839.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR001791; Laminin_C.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1827 AA; 181589 MW; AF5C71E9B38906EA CRC64;

Query Match 81.6%; Score 880; DB 13; Length 1827;
Best Local Similarity 83.2%; Pred. No. 8.9e-63;
Matches 154; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGPGFKGDMGKIDRGEIGPPGPRGEDGPGKGRGNGDPGL 60
DB 807 GIRLKGKHKGEKEDGPGFKGDMGKIDRGEIGPPGPRGEDGPGKGRGNGDPGL 866

QY 61 GPGGKGLGVPLGPGYGRQKGSIGFFPGANGKGRGTGPKGPRGQGTGPR 120
DB 867 GTVGKGLGVPLGPGYGRQKGSIGFFPGANGKGRGTGPKGPRGQGTGPR 926

QY 121 GERGPRGITKPGPKNGSGDGPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 180
DB 927 GQGRPGATGKAGKTGSGDGPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 986

QY 181 GQGRGE 185
DB 987 GQGRGE 991

RESULT 9
Q920S4 PRELIMINARY; PRT; 1053 AA.
AC Q920S4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adipocyte-specific protein 6.
GN COL5A3 OR ASP6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Tsuruga H.;
RT "Adipocyte-specific protein 6, a novel protein upregulated during
adipocyte differentiation.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB040491; BAB68504.1; -;
DR MGD; MG1:1858212; Col5a3.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF01391; Collagen; 10.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW Collagen.
SQ SEQUENCE 1053 AA; 104534 MW; C220E48BB8D1F6CC CRC64;

Query Match 68.7%; Score 741; DB 11; Length 1053;
Best Local Similarity 73.0%; Pred. No. 8.4e-52;
Matches 135; Conservative 10; Mismatches 40; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGPGFKGDMGKIDRGEIGPPGPRGEDGPGKGRGNGDPGL 60
DB 718 GNRGLQGEKGEKEDGPGFKGDMGKIDRGEIGPPGPRGEDGPGKGRGNGDPGL 777

QY 61 GPGGKGLGVPLGPGYGRQKGSIGFFPGANGKGRGTGPKGPRGQGTGPR 120
DB 778 GAAGEKGLGVPLGPGYGRQKGSIGFFPGANGKGRGTGPKGPRGQGTGPR 837

QY 121 GERGPRGITKPGPKNGSGDGPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 180
DB 838 GDRGPGATGQPGKGDVGNGSGPPGPKGLQGGPFGPKGPPGQKDGISGHP 897

QY 181 GQGRGE 185
DB 898 GQGRGE 902

RESULT 10
Q9JLI2 PRELIMINARY; PRT; 1739 AA.
AC Q9JLI2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type V alpha 3 chain.
GN COL5A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20187594; PubMed=10722718;
RA Imamura Y., Scott I.C., Greenspan D.S.;
RT "The pro-alpha3 (V) collagen chain: Complete primary structure,
expression domains in adult and developing tissues, and comparison to
the structures and expression domains of the other types V and XI
procollagen chains.";
RL J. Biol. Chem. 275:8749-8759(2000).
DR EMBL; AF176645; AAF59901.1; -;
DR MGD; MG1:1858212; Col5a3.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.

```
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 16.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; C1g helix; 5.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW Collagen.
SQ SEQUENCE 1739 AA; 171968 MW; EFDDEBA46E59984 CRC64;

Query Match 68.7%; Score 741; DB 11; Length 1739;
Best Local Similarity 73.0%; Pred. No. 1.4e-51;
Matches 135; Conservative 10; Mismatches 40; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKEGDEGPFPGKDMGIKGDRGEIGPPGPRGEGDGPBGPKRGPGNGDGPGL 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
718 GNRGLQGEKGERGEDGPFPGKDEGPKGDRGNPDPGPRGEDGPEQKGFGGLPDGEGPP 777

QY 61 GPPGKGLGVPLGPGYPRGQPKSGIGPPGFGANGKGGRTGPKGPRGQRTGPR 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
778 GAAGEKGLGVPLGPGYPRGQPKSGIGPPGFGANGKGGRTGPKGPRGQRTGPR 837

QY 121 GERGPRGITGKPGKNSGGDGPAGPGERGPNQPGTGTGFPKPGPPGPKDGLPGHP 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
838 GDRGQPGATGQPGKGDVGQNGSPGPGKEKGLPLGLQPGFPFGPKPGPPQKDGISGHP 897

QY 181 GQGE 185
DB |||||
898 GQGE 902

RESULT 11
Q9J104 PRELIMINARY; PRT; 1737 AA.
AC Q9J104;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 4 type V collagen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley;
RX MEDLINE=20428740; PubMed=10852920;
RA Chernousov M.A., Rothblum K., Tyler W.A., Stahl R.C., Carey D.J.;
RT "Schwann cells synthesize type V collagen that contains a novel alpha
RT 4 chain. molecular cloning, biochemical characterization, and high
RT affinity heparin binding of alpha4(V) collagen."
RL J. Biol. Chem. 275:28208-28215(2000).
DR EMBL; AF272661; AAF76432.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA like lec.gl.
DR InterPro; IPR008885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; C1g helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
```

```
KW Collagen.
SQ SEQUENCE 1737 AA; 171574 MW; D635D5D57481C257 CRC64;

Query Match 67.0%; Score 723; DB 11; Length 1737;
Best Local Similarity 71.9%; Pred. No. 3.9e-50;
Matches 133; Conservative 10; Mismatches 42; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKEGDEGPFPGKDMGIKGDRGEIGPPGPRGEGDGPBGPKRGPGNGDGPGL 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
716 GNRGLQGEKGERGEDGPFPGKDEGPKGDRGNPDLGPRGEDGPGGKGPEGLPDGEGPP 775

QY 61 GPPGKGLGVPLGPGYPRGQPKSGIGPPGFGANGKGGRTGPKGPRGQRTGPR 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
776 GAAGEKGLGVPLGPGYPRGQPKSGIGPPGFGANGKGGRTGPKGPRGQRTGPR 835

QY 121 GERGPRGITGKPGKNSGGDGPAGPGERGPNQPGTGTGFPKPGPPGPKDGLPGHP 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
836 GDRGQPGATGQPGKGDVGQNGSPGPGKEKGLPLGLQPGFPFGPKPGPPQKDGIPGHP 895

QY 181 GQGE 185
DB |||||
896 GQGE 900

RESULT 12
Q9AW17 PRELIMINARY; PRT; 1349 AA.
AC Q9AW17;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE S1:DZ12F11.3 (Collagen type XI alpha-2 (Fragment)).
GN COL11A2.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Skuce C.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL672176; CAD54661.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA like lec.gl.
DR InterPro; IPR008885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 12.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; C1g helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 1349 AA; 135634 MW; 1DD0432D7B63BF9C CRC64;

Query Match 49.4%; Score 533.5; DB 13; Length 1349;
Best Local Similarity 43.1%; Pred. No. 5.8e-35;
Matches 115; Conservative 14; Mismatches 55; Indels 83; Gaps 7;

QY 1 GIRGLKGTGKEGDEGPFPGKDMGIKGDRGEI----- 33
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 GIRGLKHGKGEDEGPFPGKDGKGLGVPLGPGYPRGQIKGLTGPQANGFP 448

QY 34 ---GPPGPRGED-----GPEPKRGGPN-----GDPGLG 61
```

```
Db 449 GPKGPPGPKDGLGPHGPGRGEGFQKGVGPPGVVGHPSGSETGQMGKRGHPGPPG 508
Qy 62 PGGE-----KGKLGVPGLPGYGRGPKSGISF-----PGFPGANGKGGRTGPG 106
Db 509 PPGEGQLGSPGKXEGTKDGRGPPGPKDGPGLRGPPGRLGPGTGPGLGKNGEGPAG 568
Qy 107 KPGRGQRTGPRGERGPRGIGTKPKPKNSGDDGAGPPGPRGNP-----QGPTG 160
Db 569 PPGPAGSSGRRGAGTAGTAPGPRGPPGPGPCTSGEKGVGPKGVPAGRDGIQGVG 628
Qy 161 FPGPKGPPGPKD---GLPHGPGQRG 184
Db 629 LPGPAGPGISGDDGKGEVGEPCQKG 655

RESULT 13
Q77087 PRELIMINARY; PRT; 890 AA.
AC Q77087;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Fibrillar collagen chain Papi alpha.
OS Alvinella pompejana.
OC Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Canalipalpata;
OC Terebellida; Alvinellidae; Alvinella.
ON NCBI_TaxID=6376;
RN SIGNAL 1 41
RP SEQUENCE FROM N.A.
RA Sicot F.-X., Gaill F., Exposito J.-Y., Garrone R., Deutsch J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF053538; AAC35289.2; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clq helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Clq helix; 9.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Collagen.
SQ SEQUENCE 890 AA; 88137 MW; 52CC756FA70CA90C CRC64;

Query Match 47.2%; Score 509.5; DB 5; Length 890;
Best Local Similarity 47.5%; Pred. No. 3.3e-33;
Matches 106; Conservative 14; Mismatches 64; Indels 39; Gaps 5;

Qy 1 GIRLGLTKGKEDGPPGPKGDM---GIKGDREIGPPGPRGEDGP---EGPKGRGGPN 54
Db 355 GTPGMDTKGERGEDGPPGLPGTGPPOGGERGVGLFGERGEPGPGAPGRGPGAP 414
Qy 55 GDRPGLPPGEEKLGVPLGVPYGRGQPKSGTGFPFGANGKGGRTGPKGP-----110
Db 415 GEPGDTGPRGERSGSPGAPGAPGAAGPEGPKDGLPLAGRPDKGPPGPPGGLA 474
Qy 111 -----RGQRTGPRGERGRTGKPKGNSGGGDPAGPP-----GERGNGQP 156
Db 475 GARGMQLRPPGPTGPGGERGERGFGAPGVGRVPGVPAGAPQOQSKGRGAAGPK 534
Qy 157 GPTGFP-----GPKGPPGPKDGLPHGPGQRG 184
Db 535 GDKGWFPGMPLQGMQCTPGERGEPGPPGPGADGMKRGEGSRG 577

RESULT 14
Q81ZC6 PRELIMINARY; PRT; 1860 AA.
ID Q81ZC6
AC Q81ZC6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
```

```
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen XXVII proalpha 1 chain precursor.
GN COL27A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pace J.M., Corrado M., Missero C., Byers P.H.;
RT TISSUE=Cartilage;
RT "Identification, characterization and expression analysis of a new
RT fibrillar collagen gene, COL27A1.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY149237; AAN41263.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; ConA like lec_gl.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 15.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR ProDom; PD00038; COLFI; 1.
DR SMART; SM00210; TSPN; 1.
KW Signal.
FT SIGNAL 1 41
FT CHAIN 625 1621
SQ SEQUENCE 1860 AA; 186891 MW; 5F8CDFAF4B6014EC CRC64;

Query Match 46.9%; Score 506.5; DB 4; Length 1860;
Best Local Similarity 36.4%; Pred. No. 1.2e-32;
Matches 121; Conservative 20; Mismatches 44; Indels 147; Gaps 13;

Qy 1 GIRLGLTK-----KGEKEDGPPGPKGDMGINKDRGE-----32
Db 838 GMKGLMGVSGEPLGKDGKGEQGVGSDPGDFGQDKGSQGLPGFPGARGKPGPLGKVGDK 897
Qy 33 -----IGPPGPR-----GEDGPE-----GPKGRGPGNDGPPGLGP-----62
Db 898 GSIGFPGFPGEGFPGDGIIPGPDNGPEGMKGPARGLPGPRQLGPEGDEGPMGPGAP 957
Qy 63 -----PGEKGLGVPL-----PGYPRGQPKSGTGFPGF-----PGANGK-----99
Db 958 GLEGQGRKGFPRPGLDGVKGEPPGDPGRPPVGEQGFGLVGEPLVGEKGRGMM 1017
Qy 100 -----GGRGTGPKGP-----RGQRTGPKGP-----123
Db 1018 GPPGVPGPKSGMHPGMPGCMGTGPGEPGQPGSGRPGMRGAKGRRGPRGPDGAGEQ 1077
Qy 124 GPRGITGPKPKNSG-----GD-----GPAGP-----PGRGNGQP 156
Db 1078 GSRGLKGPGRPCRPGRPCQGVAGRGHLGSRGFPFGIPGSPGPGTKGLPGPGPGQ 1137
Qy 157 GPTGFP---GPKGPPGPKDGLPHGPGQRG 185
Db 1138 GPIGPPGEMGPKGPPGAVGEPGLPGEAGMKGD 1169

RESULT 15
Q9NUB7 PRELIMINARY; PRT; 886 AA.
ID Q9NUB7
AC Q9NUB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DA2A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
GN COL4A5.
OS Homo sapiens (Human).
```

Job time : 22.8274 secs